DOCUMENT RESUME

ED 325 360 SE 051 696

AUTHOR Westbury, Ian; Travers, Kenneth
TITLE Second International Mathematics Study.
INSTITUTION Illinois Univ., Urbana. Coll. of Education.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE 90
CONTRACT SPA-87-51425
NOTE 315p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC13 Plus Postage.
DESCRIPTORS *College Mathematics; *Comparative Education; Databases; Data Interpretation; Developing Nations; *Elementary School Mathematics; Elementary Secondary Education; Higher Education; International Cooperation; *International Studies; *Mathematics Achievement; Mathematics Curriculum; Mathematics Education; Psychometrics; *Secondary School Mathematics; Sex Differences
IDENTIFIERS International Assn Evaluation Educ Achievement; *Second International Mathematics Study

ABSTRACT

The papers collected in this volume are intended both to present a summary 'history' and description of the International Association for the Evaluation of Educational Achievement (IEA) Second International Mathematics Study (SIMS) and to illustrate a variety of approaches to the analysis of the data that emerged from SIMS. The papers are intended to encourage others to explore the SIMS database for both national and comparative studies on the teaching and learning of mathematics. Included are: (1) an overview of the IEA study; (2) a list and review of doctoral theses using SIMS data; (3) a summary report of content in college algebra; (4) a gender difference report; (5) a model of school effectiveness in a developing country; (6) a paper n psychometrics; (7) non-cognitive data; and (8) a report on state control of the curriculum. Appended are a technical report on college algebra and a list of the national and international reports on SIMS. (KR)

SECOND INTERNATIONAL MATHEMATICS STUDY

Studies

Ian Westbury
Kenneth Travers

COLLEGE OF EDUCATION
UNIVERSITY OF ILLINOIS AT URBANA - CHAMPAIGN
1990

This material is based upon work supported by the National Science Foundation under Grant No. SPA-87 51425. The Government has certain rights in this material.

Table of Contents

INTRODUCTION

The papers collected in this volume are intended both to present summary 'history' and description of the IEA Second International Mathematics Study and to illustrate a variety of approaches to the analysis of the data that emerged from SIMS. The initial results from SIMS have been reported in both the series of national reports outlined in the Appendix to this report and in the technical and published reports describing the results of the international analyses. However, as numerous observers of SIMS have commented and the papers in this report amply illustrate, these initial analyses--by members of the core working groups closely associated with the study throughout its now long history--have barely scratched the surface of the data that SIMS collected and have not explored the variety of questions and the analytical approaches and methods that the SIMS database can support. The papers collected here are intended to encourage others to explore this rich database for both national and comparative studies on the teaching and learning of mathematics.

Most of the papers in this report were initially presented at a seminar on Secondary Analysis of the SIMS Database held at the University of Illinois at Urbana-Champaign in January 1989. Subsequently we became aware of the work of David Baker and David Stevenson of the Catholic University of America and took advantage of their willingness to share their exciting research with the SIMS 'community' by incorporating early versions of two of their papers in the Report. In addition this report includes abstracts and an initial review, prepared by Leigh Burstein of the University of California, Los Angeles, of some of the recently-completed U.S. dissertations that have used the SIMS data.

The University of Illinois at Urbana-Champaign seminar on Secondary Analysis of the SIMS Database and this Report are part of a larger project, the SIMS Database Enhancement Project, which has as its major tasks the preparation of the public-use database which might support further secondary analysis of the SIMS data, the training of researchers in the use of both the SIMS data and the database, and the encouragement of secondary analysis of SIMS data. The SIMS Database Enhancement Project is supported by a grant from the United States National Science Foundation (Grant No. NSF SPA 87-51425).

Ian Westbury
Kenneth J. Travers

ACKNOWLEDGEMENTS

This Report could not have been completed without the willing assistance of the authors of each of the papers included in the seminar on Secondary Analysis of the SIMS Data. We must acknowledge the willingness of Marlaine Lockheed of the World Bank to come to Champaign-Urbana in January--when she might have been in warmer places--to report on her on-going analyses of the data from the low-income systems which participated in SIMS. As noted above, we are also particularly grateful for the willingness of Deborah Perkins Jones, David Baker and David Stevenson of Catholic University to allow us to incorporate reports of their work in this volume. Edward Kifer of the University of Kentucky and Richard Wolfe of the Ontario Institute for Studies in Education, Toronto, also willingly permitted us to include their descriptive overview of SIMS; this paper was written as part of unpublished report prepared for the U.S. National Center for Education Statistics. In addition, we must acknowledge the assistance of Denny Arvola, Letitia Klatt, and Myrna Craig of the College of Education at the University of Illinois at Urbana-Champaign in managing the complex administration of the seminar which was the starting point for this Report. Later Leigh Little, Pam Bahal, and Charles Ramer of UIUC did substantial work in perfecting the final copy for this technical report. Of course, the original conference that was the source of this Report and the SIMS Data-base Enhancement Project would not have been possible without the support of the National Science Foundation and the on-going assistance of Richard Berry, Project Officer.

Ian Westbury

AN OVERVIEW OF IEA STUDY

R. A. Garden
Director Research and Statistics
New Zealand Department of Education

Introduction

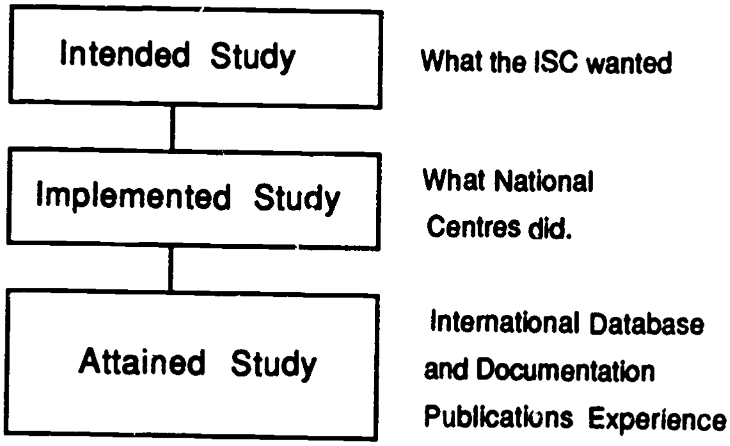
In the course of the Second International Mathematics Study (SIMS) conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), data was obtained from approximately 3,900 schools, 6,200 teachers, and 124,000 students in more than 20 education systems around the world. This discussion will not deal with detail about the aims or conduct of the study - these can be read in a number of study's publications, the set of five Bulletins for example. Nor will it acknowledge the contributions to the study of a large number of individuals. To the extent that recognition of the dedication, perserverance, and special skills contributed by so many can be made by a few words in a publication this also has been done elsewhere. My intention to give an indication of some of the influences and constraints that determined the nature of the databank which is a major outcome of the study. It is to be hoped that users of the databank who find that their favourite variable was not included in the study, or who discover flaws in the data, or shortcomings in the documentation, will think about the difficulties associated with the scale of the project, its multinational nature, and the miniscule levels of funding for key phases of the project--and refrain from rushing into print with trenchant criticisms.

Data collection, preparation and analysis took place in the first half of this decade and it is now beginning to be easier to place the various SIMS actors and their actions, and all the activities of the study, into some sort of perspective. During the study it was the negative aspects which dominated our lives - the National Research Coordinators (NRCs) who did not follow instructions, the postal delays, the misunderstandings, the unreadable data tapes, the mis-coded data, and so on. Now distance in time is beginning to lend a degree of enchantment to the view, but in order to give some shape to the overview it helps to adopt a suitable framework.

There are several possibilities. The first which commended itself was a framework based on the life-cycle metaphor, where progress would be discussed in terms of conception of the study (perhaps there was even a seduction phase), its birth (not without subsequent post-natal depression), toddlerhood (poverty-stricken but filled with hope), adolescence (sturm und drang), maturity (great responsibility but still no money), and old age (now some money but too late to enjoy it). There doesn't seem to have been an identifiable death scene, but some of us are certainly being haunted, and this meeting might well be the platform for resurrection. A battle metaphor also suggested itself, with declaration, arming and training phases and so on. But to sustain this metaphor one would have to talk of skirmishes, conflicts, tensions between generals in the rear and lieutenants down the line, perhaps even some putting of blind eyes to the telescope, and of course all this would be quite inappropriate. Or would it!

The solution to the problem of how best to map an overview emerged as I re-read the B·lletins produced throughout the study. Those familiar with the study documents will recognise the source of the model at once, and those who have been involved with the study throughout its planning, development and execution will acknowledge its aptness. Any overview of SIMS would have to recognise the existence of the study phases shown in Figure 1. There are important differences between what was intended and what happened, and between what was sought and what was captured. Without an awareness of the causes of these differences misinterpretation of results of analyses of the data is a strong possibility.

Figure 1.



Framework for the Overview

The shape of this overview, then, is similar to that by which SIMS participants came to view the curriculum. The model for the study is, I believe, adaptable to a range of diverse human activities and organisational processes, but it should be noted that the interpretation I will share with you is my interpretation. Other people involved in the study would doubtless have different interpretations. One of the unforgettable lessons administrators of cooperative international studies learn is that amongst people from different social, cultural, political backgrounds it is very hard to find a common perception of many of the things whose meanings we, as individuals, take for granted within our own socio-cultural environment.

Another factor to be taken into consideration is that although I was involved in the study from fairly early in its evolution, I did not become International Coordinator until 1980. Roy Phillipps, the first International Coordinator (and at that time a member of the IEA Standing Committee), and Ken Travers, Chairman of the International Steering Committee (ISC) throughout the study may not share all my perceptions. My view of early attempts at fundraising and initial planning for SIMS was very much a worm's eye view. Figure 2, then, provides the framework for my talk.

It should also be noted that my view was from the International Centre in the Department of Education in Wellington, New Zealand. Work related to the longitudinal component of SIMS, and to the construction of the relevant sections of the databank, was carried out at the study Centre at the University of Illinois at Urbana-Champaign with support from Richard Wolfe and others at the Ontario Institute for Studies in Education in Toronto. However, the challenges associated with processing IEA data appear to be independent of geographic location.

Study Antecedents

Features such as the kind of data that are in the SIMS databank, the way they are arranged, their characteristics and quality are, in part, the result of events which occurred before the study was conceived. Earlier IEA studies, the First International Mathematics Survey (FIMS) and the Six Subject Survey, had seen the development of a research design and methodology which had 'worked'. Leading researchers from several countries had been involved in the cooperative development and execution of the previous studies and now formed the nucleus of an extensive international community of comparative researchers. SIMS thus became the latest 'baby' of the IEA family and inevitably manifested a strong genetic blueprint. This carried many benefits - and a few disadvantages - but the point to note is that the study is unmistakably an IEA study.

Because this was the second IEA mathematics study there was an especially strong pattern in existence and, in fact, some participants initially saw the second study as being a replication of the first. But although there was strong interest in being able to make comparisons over time, rapid growth in the mathematics education knowledge-base, an expanding set of techniques of analysis, and competing views of how research should be carried out quickly reduced the importance of this aim for the study.

SIMS was intended to have a much stronger emphasis on mathematics education than had FIMS, where mathematics tended to be treated as a surrogate for school achievement in general. To this end, the ISC, advisers and National Committees included strong representations of respected mathematics educators. The questions they wished the study to address were diverse, and many came from a different domain from those examined in prior IEA studies. Furthermore, many of the questions could not easily be addressed by traditional IEA survey methods. It is not surprising, given differing expectations of the study, that from time to time 'father' IEA Standing Committee did not always see eye to eye with 'mother' ISC over how the SIMS 'baby' should be reared.

At national (or system) level, the character and history of the institution which houses the IEA National Centre partly determine how much influence that centre will have on international instruments and manuals, as well as on the assiduousness with which international instructions about data collection and preparation are followed. Previous institutional experience of participation in an IEA study, or of other large scale survey work, can be expected to contribute to a more complete and error-free national data set. For the same reasons, the experience, research ability, and other personal qualities of the National Research Coordinator are reflected in the study outcomes. However, it should be noted that in SIMS some excellent data sets came from systems in which relevant experience was not great. An ability (and willingness) to follow instructions to the letter was the main pre- requisite required.

Study Contexts

Past IEA studies were an important antecedent to SIMS. The individuals who had played leading roles in these studies and who occupied influential positions in IEA, and in the research community generally, when SIMS began constituted an important contextual factor.

Their experience, knowledge, beliefs and attitudes had a considerable influence on the conduct of the study. A second contextual factor was what might be called the "IEA ethos". Early IEA studies were pioneering ventures which, because of the way IEA had come into being, attracted researchers for whom rewards such as the intellectual challenge, intrinsic interest, and stimulation of participating cooperatively in an international venture were enough. Many, perhaps most, of those who play leading roles in IEA studies are prepared to sacrifice a vast amount of their time to engage in very difficult tasks for which they receive little, if any, financial recompense. But by the time SIMS was underway a number of researchers simply could not afford this sort of financial sacrifice. Finding ways of ensuring that amongst the best available consultants and advisers were able to make key contributions to the study utilised a good deal of the energy and challenged the imagination of study administrators.

Funding, or rather the lack of it, was undoubtedly the greatest handicap faced by those planning and executing the study. Policies of funding agencies had changed substantially since earlier IEA studies. Many were targetting funding locally, or at specific reserrch fields, and were just not "in the market" to provide funding for international research. Those few funding international research were not necessarily interested in secondary school mathematics research.

For a large part of the duration of the study there was just enough support to allow progress to be made. Maintaining the commitment of research teams to difficult tasks when there is no guarantee that the manuals and instruments they are producing will ever be used is not easy, and relying only on correspondence (in English) for communication with researchers from a wide range of nations is definitely not recommended. Inability to fund meetings of National Research Coordinators early in the planning phase of the study resulted in a certain amount of "undoing and redoing" of work. In general problems of this sort were solved satisfactorily, but there are one or two places where the repairs are obvious. A case in point is the sets of items added late to the cognitive instruments in an attempt to meet criticisms that the curricula of certain European systems were not adequately represented.

Despite the bleakness of the general funding picture at the start of the study, there were people in some of the funding agencies who sensed the potential of the study and were able to furnish assistance. As the likely outcomes of the study became more apparent further funding was able to be obtained, but even at this stage this was a far from easy task for IEA advocates within the agencies. Those associated with SIMS are very grateful for the good work they did. Some of these agencies made further substantial contributions to the study through the expertise of professionals on their staffs or through the expertise of researchers they recommended and encouraged to assist with aspects of the study.

At national level, the levels of funding and resources available were also of crucial importance. In many National Centres researchers carried out their duties as National Research Coordinator in addition to a substantial workload from national or local projects. Furthermore, it was common for them to have to give less priority to SIMS than to other projects they were working on. In the face of difficulties of this sort the NRCs did a remarkable job. It is not surprising that there were, from time to time, problems in meeting deadlines or in meeting specifications in the provision of data.

The quality of IEA studies depends as much on the NRCs as on any other group. It is they who are responsible for translating the wishes of the ISC into action in cultures and national milieux quite different from those of people who exerted the major influences on design of the study. In the phase in which instruments and manuals are being negotiated they must convey the spirit and intent of the study to their national committee members, and represent the wishes of their national committees to the ISC.

The success with which SIMS NRCs could meet these demands depended on their having credibility with mathematics teachers as well as with the research community. The management challenges on an IEA study are as significant as the research challenges - and their importance is sometimes overlooked.

We need to remember then, as we examine and manipulate IEA data, that the contexts in which the study was designed and executed had a lot to do with the nature of study outcomes. Those carrying out secondary analysis would be well advised to bear this in mind and to make a real effort to "get a feel" for the contextual factors in those countries for which interpretation of analyses will be made.

The Intended Study

Nothing is ever simple in large-scale studies where design and methods are negotiated by participants from diverse cultural backgrounds. To refer to SIMS as a single study disguises the scope and complexity of what was really a collection of studies. To begin with there were two target populations:

Population A: All students in the grade (year level) where the majority have attained the age of 13.00 to 13.11 years by the middle of the school year.

Population B: All students who were in the normally accepted terminal grade of the secondary education system and who were studying mathematics as a substantial part of their academic programme.

At each grade level systems had the choice of administering the full study (i.e. a longitudinal study which included pretest and post-test and collection of classroom process and teacher behaviour data) or doing a "reduced" study (cross-sectional with post-test as the only cognitive measure and without classroom process data.) The participating systems, with the populations tested and the version of the study are shown in Table 1. Ontario, British Columbia and the United States were the only systems to undertake a longitudinal study for Population B.

Table 1 Participating Systems

System	Population(s)	Study* (Pop A)	System	Population(s)	Study (Pop A)
Belgium (Flemish)	A & B	L	Luxembourg	A	C
Belgium (French)	A & B	C	The Netherlands	A	C
British Columbia	A & B	L	New Zealand	A & B	L
England & Wales	A & B	C	Nigeria	A	C
Finland	A & B	C	Ontario	A & B	L
France	A	L	Scotland	A & B	C
Hong Kong	A & B	C	Swaziland	A	C
Hungary	A & B	C	Sweden	A & B	C
Israel	A & B	C	Thailand	A & B	L
Japan	A & B	L	USA	A & B	L

*Longitudinal, Cross-sectional

The variations resulted in large part from the outcomes that national centres saw as having most value for their systems at that time. For many, a comparison of mathematics achievement between their system and other systems was the most important outcome desired. Half of the systems were interested in obtaining an indication of whether mean student performance in their system had improved or declined since FIMS. The group of systems comparisons had as their prime interest identification of variables that could be manipulated to improve mathematics achievement.

SIMS was intended by the ISC to differ from FIMS in another important way. As well as the introduction of the longitudinal study, there was a thorough analysis of the curricula of participating systems. This was seen not only as having intrinsic interest, but also as a way of illuminating the results of the cognitive tests.

Notwithstanding the variations built into the design, the "intended study" as far as the ISC were concerned involved all systems completing all tasks and instruments in the components of the study they had elected to participate in, following the detailed instructions in memoranda and manuals to the letter, and hence producing flawless data sets or, at worst, data sets with all deviations and omissions carefully documented.

The Implemented Study

Probably no national centre administered the study exactly as intended by the ISC. Nor could it realistically be expected that they would. Each member institution operates under its own constraints and within its own national culture.

Different National Committees develop different aspects of IEA study's as priority areas and, if resources are scarce, may delete low priority questions or instruments. In a few cases questionnaire items were changed, or mistranslated. In a few cases some questionnaire responses were precoded (e.g. Language of the Home was assumed to be Japanese for all students in Japan, and in a couple of systems periods lengths and days in the school year were assumed to be constant across schools).

Arriving at definitions of target populations and constructing sampling manuals which can be implemented in an identical fashion and which have the same effects and results in all systems is just not possible. But the sampling manuals used in SIMS were based on the experience of past IEA studies combined with the wisdom of acknowledged experts in sampling. A process which involved NRC's in comment, negotiation with the SIMS Sampling Committee, and approval of sampling plans by a sampling referee was designed to minimise sampling errors and to make outcome measures comparable across systems. National samples which fell short of enabling these ideals to be fully attained did so for a variety of reasons as documented in the Sampling Report for the study. But as I asserted in that report, even for the least satisfactory samples, enough is known about them for some important conclusions to be drawn with reasonable confidence. However, there are data for variables in some systems which should be interpreted with great caution and are better not included in multi-variate analyses.

The data collection phase was generally well executed. Where response rates were not as good as were hoped for this was not through inadequacies in the manuals or other advice sent from the International Centre in Wellington, nor was it through lack of diligence on the part of NRCs. In the worst case the Nigerian NRC was unable to get to several provinces because of extensive flooding, so the population definition for Nigeria was changed to take account of this.

The least well executed part of the study at system level was in the preparation of data for shipping to the International centre. With the wisdom of hindsight it is now clear that in some centres this resulted from lack of experience in handling datasets of the size of those in SIMS, i.e., several thousand cases with several lengthy records for each case. Insufficient clerical provision had been made for checking and coding and several NRCs experienced weeks of tedious work. There was also a considerable range of expertise amongst NRCs in computer-related data preparation (but it should not be assumed that poor tapes were received only from less experienced, or good tapes only from more experienced national centres. Experience helps, but the ability to follow instructions was just as crucial). The outcome was that the data received at the International Centre from national centres posed a series of challenges.

The first of these challenges was to read the tape - not always possible for the first tape received. The second was to decipher what was on the tape. The third was to relate what was on the tape to what was expected to be on the tape. The fourth was to caress (or sometimes bash) the data set so as to get it into the required format, remove out of range values and impossible outliers, without unnecessarily losing one "good" data element.

Working through these stages was a long process. Where data could not be read from the tape, the national centre had to be asked for a new tape. Even when tapes were read, new tapes had to be asked for in some cases, but this was not common because great efforts were made to get the data into shape at the International Centre. This was judged to be likely to take less time than sending the faulty tape back to the national centre and waiting for it to be dealt with, especially as national centres tended to have used all their SIMS funding by that stage. When the data was readable and correctly formatted in was checked for out-of-range data, outliers, accuracy of details of modifications and deletions supplied by each national centre, and any unexplained anomalies. The reasons for these were often obvious and appropriate editing could be done at the International Centre. Other anomalies were able to be corrected at the International Centre after considerable detective work. NRCs were sent frequency outputs for each item in their data sets and asked to check them. They were asked to check and approve changes made at the International Centre and to explain any anomalies which the International Centre had been unable to resolve.

There could thus be several exchanges of correspondence between the International Centre and a national centre (especially as some NRCs did not respond to correspondence for some time). If the study were being conducted now, with the availability of E-mail and Fax, this process would be drastically shortened. As it was, it all took a long time. The alternative was, in my view, the loss of a great deal of data, and possibly dropping some systems from the study.

The Attained Study

What do SIMS veterans have to show for their work up to now? Already there is a substantial list of publications associated with the study. Two of the three volumes of the international report are available and the third will be published in the near future. Other substantial publications for an international audience are planned.

Articles have appeared in international and national journals, and short research briefs have been prepared for national audiences. But perhaps the greatest impact of the study has been made through the national reports produced by national centres. It is these, written from the perspective of the participating system for a home audience, that elicited responses from politicians, educational administrators, mathematics educators and teachers who were made aware of a need to, or a way of, improving mathematics education in their schools.

Not to be underestimated either, are the long term effects on a system's mathematics education and research communities of having participated in SIMS. Researchers learned new techniques and refined old ones, mathematics educators were introduced to new ways of looking at their field, and teachers in many of the systems in the class-room processes component of the study remarked that participation had been excellent in- service training in mathematics education.

But the most important outcome of SIMS could yet prove to be the resource which will be under discussion during the next few days. Although there has been a considerable amount published from SIMS data, the surface has scarcely been scratched. There can be few databanks as extensive and as complex which have had the same amount of careful work put into them to keep the data as complete as possible, to provide extensive explanatory documentation, and to make the data accessible, as this one.

Discussion

This narrative does not amount to much more than a rather sketchy outline of what was a decade's endeavour involving many people. It focusses on those features of the study that led to its most noticeable outcomes. Another overview might have traced the changes which took place in the intended outcomes and study procedures as more systems committed themselves (late) to participation, or as understanding of the standpoints of already participating systems grew. Changes in emphases within mathematics education, and education generally, also gave rise to new emphases as planning progressed. For example, early in the planning the use of calculators in mathematics, applications in mathematics, and minimal competency policies were projected as being major features of the proposed study. These topics eventually receded into the background, but ERIC publications featuring discussions of the then current views and activities in these areas from each of a wide range of countries were among the important, but less visible outcomes of the study.

. The real guide to the degree of success achieved by SIMS will lie in answers to the questions: Did the audiences which the ISC targetted "receive" useful messages from

SIMS? Has appropriate action resulted? Does the SIMS experience add to our knowledge about research/mathematics education and suggest lines of inquiry for future study?

For the first pair of questions the answer would be a qualified yes. We know that SIMS has led to action designed to improve mathematics achievement from policymakers, mathematics educators and teachers in some of the participating systems. The qualification is because we do not have information from every system about the impact of their national report, and because when action is taken it is usually not on the scale that the research suggests is needed. Educational administrators, and often teachers, tend to impart a "regression" effect towards the status quo.

The answer to the final question is an unqualified yes even at this relatively early state and it is certain that knowledge about mathematics teaching and learning, and about research into these, will increase if the databank is widely utilised. One would hope that alternative models, both of learning and of analysis, will be tested. There is scope for rethinking of which are the key variables and how these variables might be constructed. It would be of interest to replace achievement as the dependent variable by Implemented Coverage, or Teacher Expectation of Success. Replacing mean class achievement with percent of class reaching a given mastery level as dependent variable might also lead to some interesting results. The possibilities are almost limitless.

Another field which might be explore via the SIMS data is that of educational indicators. Many education systems throughout the world are seeking measures of the "health" of their systems. A large OECD exercise is currently underway in this field. Indices constructed from coverage (opportunity-to-learn) provide measures of "conformity" (between what was expected of teachers and what they did in teaching mathematics), and of "efficiency". Other IEA variables, yield for instance, suggest themselves as indicators. Supplementing IEA data with up-to-date financial information would give rise to a further set of indicators.

Conclusion

Already there is talk of a Third International Mathematics Study, (which demonstrates again the healing power of time). A very few years ago the mere suggestion of going through it all again would have brought on nightmares. But there should be a TIMS.

Shortcomings in this sort of study are inevitable, but important difficulties experienced with SIMS should now be able to be minimised. We know which procedures work and which do not, which national centres need extra support and the kinds of support they need. We know which variables worked, how to improve the measurement of some of those which did not do so well, and which variables were missing.

Improvements in communication (E-mail and Fax) would shorten a third study by years, assuming reasonable levels of funding. Without funding adequate to provide for cooperative planning and design of the study with NRCs, execute the study, and provide for production of international reports, all bets would be off. I would not like to see the general nature of IEA studies change, but if we cannot have executives working full-time on the TIMS to ensure that it runs to schedule it would be better to abandon the field to the "fast test" experts. (Reading the SIMS Bulletins will reveal that the study schedule was a systematic variable with substantial variance. Arguments that, like fine wine, the SIMS data would improve with age did not win approval.)

Perhaps IEA should update its fund winning methods. SIMS could just as easily have stood for the Steinlager International Mathematics Study. New Zealand's results could well have driven mathematics teachers to seek solace in that fine product.

In almost all substantial research projects it can be claimed that the data is grossly under-exploited. Major efforts have been made to preserve the SIMS data in a form in which it is readily accessible and interpretable to researchers for further analysis. All of those who were involved in the SIMS enterprise will be delighted that the data will continue to be used towards the improvement of mathematics teaching and learning. Every effort must be made to see that researchers in many countries make full use of this resource.

The Studies

Edward Kifer
University of Kentucky

Richard Wolfe
Ontario Institute for Studies in Education

Introduction

This chapter contains descriptions of a few major differences among the countries that took part in the longitudinal portion of the Second IEA Mathematics Study. The first section gives size, population and geographic information. The second provides a brief summary of the structure of the school systems. Third, there is a brief synopsis of the characteristics of the study conducted in each of the countries. finally, there is a general overview of the curricula of the countries. This and the second chapter are meant to set a context in which the results can be interpreted and responded to.

The Countries

This volume focusses on teachers, students and classrooms and how students change during a year of schooling. The international nature of the study, however, serves as a constant reminder that education is essentially a social and cultural phenomena. Students do change while in schools and some of that change is because they are in schools. But all of what they learn is embedded in a context defined by differences in values, geography, wealth, tradition and any of a variety of variables that can be summed up rather easily. These are different countries.

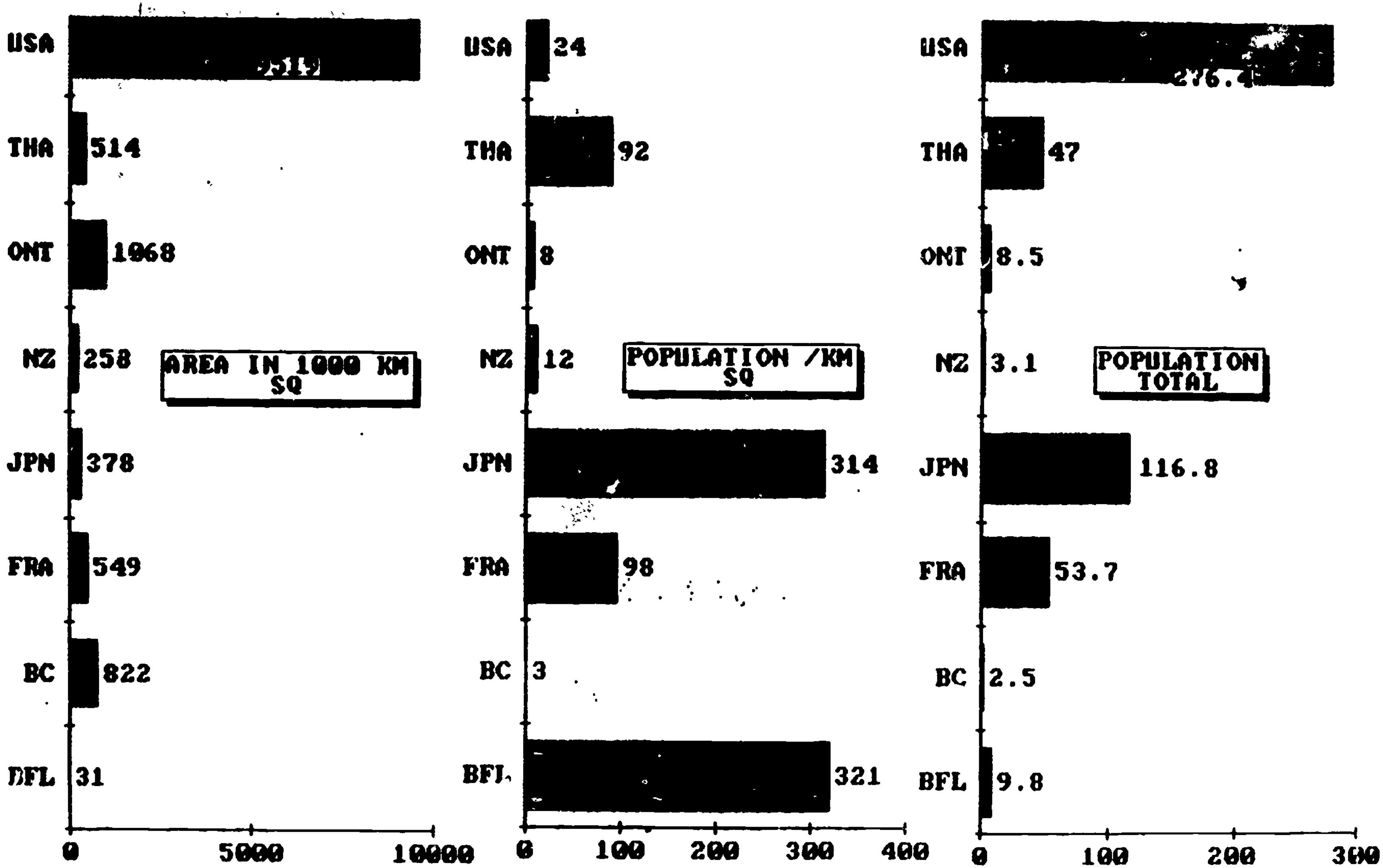
There is a story[1] told by an Australian journalist about the Japanese sense of "wah" (cooperation, harmony and balance) and how it pervades virtually every facet of social and economic life. While living outside of Tokyo, he engages in typical activities, one of which is buying gasoline for his automobile. The station he frequents charges a few more yen per liter than one slightly farther away from his home. He is, of course, free to change where he buys gasoline. Should he do that, however, the owner who now has the journalist's business has an obligation (wah) to the previous owner to compensate him for the loss of a customer. The amount and type of compensation is determined through long and involved bargaining within a context of unwritten but complicated rules. The journalist, on the other hand, has an interest in staying with the first owner since it is also understood that should demand evaporate for his stories about Japan, gasoline would be available and a tab kept for any reasonable amount of time. After all, owning a gasoline station is much more (wah) than merely selling gasoline to make money.

It would be possible, one imagines, to give anecdotes such as this for each of the countries in the survey. Each maintains traditions and ways of operating that would be "foreign" to others. Tacit understandings, language differences, customs, traditions and other "cultural" variables are not measured in this survey. But there is no doubt that they, just as some of the background conditions described below, influence profoundly the experiences of children in schools.

Figure 1 shows in terms of population and area[2] other differences between these countries. The USA, for example, is an area 300 times larger than Belgium Flemish and in population 100 times larger than British Columbia. The population density of Belgium Flemish and Japan are 100 times greater than that of British Columbia. Though small, Belgium Flemish is very heavily populated; though relatively small, New Zealand is rather sparsely populated. How and to what extent these factors influence educational processes are matters for healthy speculation. To posit that such factors do not influence directly or indirectly schools and schooling would be folly

Figure 2 contains demographic and educational factors[3] that differ between these countries. What is estimated to be the wealthiest country, British Columbia, has a per capita income 200 times larger than that of Thailand, the poorest country. One hundred percent of the students in Japan responded that the language of the school and their home is always the same while only 16% of students in Belgium Flemish gave that response. Students in the United States are exposed to 50% more mathematics instruction in a school year than are students in Japan. Still, the estimated 150 hours per year for the USA represents at best 15% of the time that students spend in school and, if one imagined an intensive mathematics course that lasted 6 hours per day seven days per week, it would be only about 3 1/2 weeks of the year that they were in a setting where mathematics is taught. It is a small part of a child's life that is devoted to receiving mathematics instruction.

Enrollment figures for the eight countries show equally dramatic differences. Since 1965, about the time of the first IEA mathematics study, the seven developed countries have had rather stable and sometimes declining total enrollments at the primary levels of schooling. Thailand during that period increased its primary school enrollment by almost 3 million students or an increase of almost 75%.[4] Its lower secondary schools increased six-fold, from a 1965 total of about 250,000 to a 1980 total of 1,500,000. The developed countries found another way to expand schooling--make more of it almost mandatory. France in those 15 years changed from a university enrollment of 400,000 to 1,000,000. Japan almost tripled the number of students in higher education from a 1965 total of 800,000. Both community college and university enrollments expanded rapidly in British Columbia and Ontario. So there has been expansion of



AREA IN 1000 KM SQ
USA
THA 514
ONT 1068
NZ 258
JPN 378
FRA 549
BC 822
BFL 31
0
5000
10000
POPULATION /KM SQ
USA 24
THA 92
ONT 8
NZ 12
JPN 314
FRA 98
BC 3
BFL 321
0
100
200
300
400
POPULATION TOTAL
USA
THA 47
ONT 8.5
NZ 3.1
JPN 116.8
FRA 53.7
BC 2.5
BFL 9.8
0
100
200
300

14a

TABLE 1

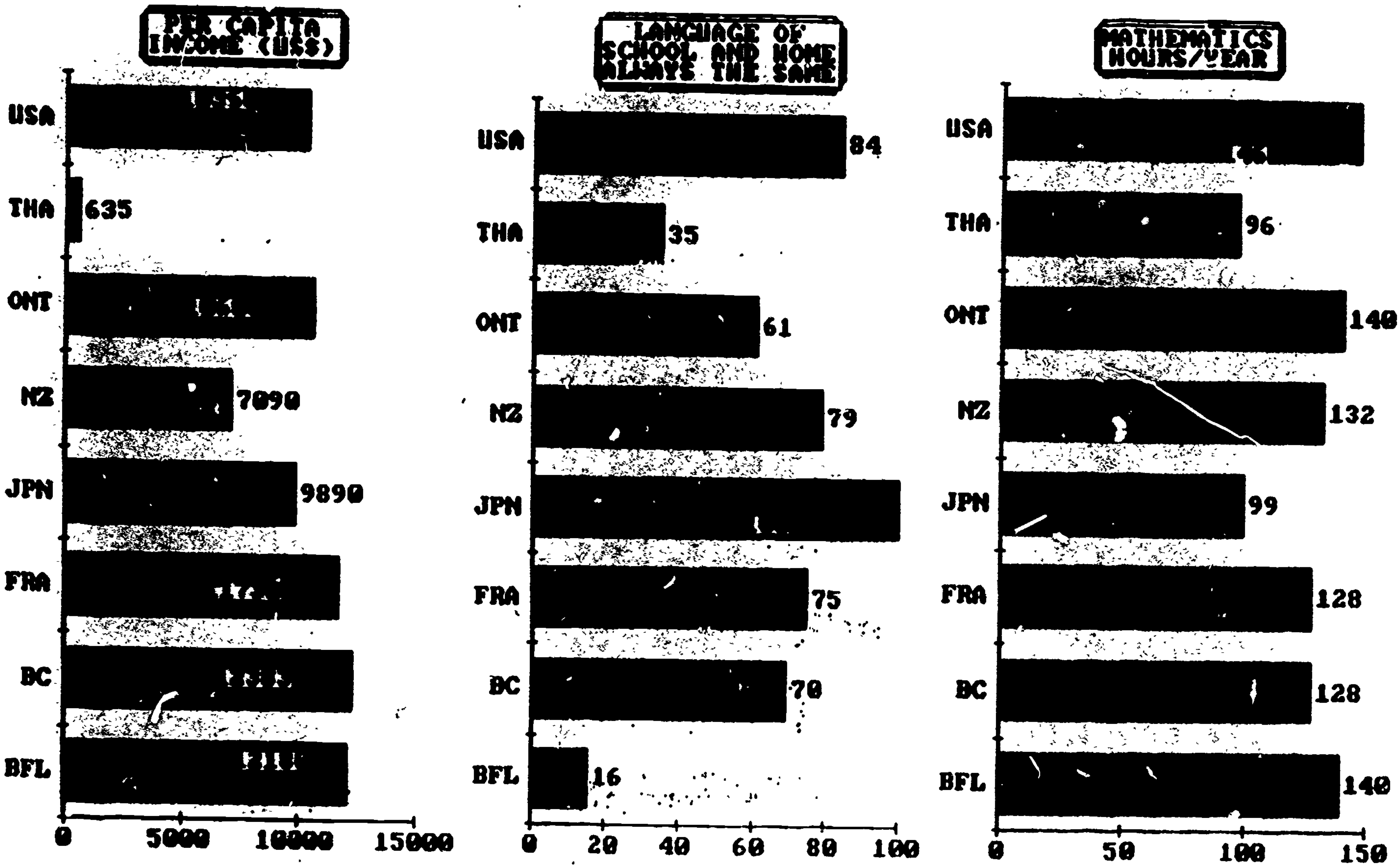

PER CAPITA INCOME (US$)
USA
THA
635
ONT
NZ
7090
JPN
9890
FRA
BC
BFL
0
5000
10000
15000
LANGUAGE OF SCHOOL AND HOME ALWAYS THE SAME
USA
84
THA
35
ONT
61
NZ
79
JPN
FRA
75
BC
70
BFL
16
0
20
40
60
80
100
MATHEMATICS HOURS/YEAR
USA
THA
96
ONT
140
NZ
132
JPN
99
FRA
128
BC
128
BFL
140
0
50
100
150

TABLE 2.

schooling in all countries. The major differences is at what level of a school system the enrollments grew.

Despite those rather obvious differences, the countries apparently share similar views of the power and importance of schooling. Children begin to go to school in each country at either age 5 or 6 and end no earlier than age 15 or 16. A trend appears, however, to be in the direction of extending both the duration and universality of expected time in schools. A 1974 reform act in Belgium Flemish made at least halftime attendance compulsory until age 18 while students in Ontario can attend publicly supported schools at age 4. Formal schooling (from cradle to grave?) is expanding, with the developed countries increasing participation in schooling beyond the secondary level and the developing ones making primary and secondary education universal.

The Structure of the Schools

Those who are familiar with U.S. schools know how difficult it is to describe how they are organized. The sample, for instance, of students in this study comes from the eighth grade. Is that the end of elementary school, the second year of junior high school or the end of a middle school that separates elementary and secondary schooling? Other countries have equally ambiguous organizational structures so what follows is a description of typical patterns rather than a presentation of these school organizations in all of their complexity.

Belgium Flemish

Pre-schools are available to children ages 2 1/2 to 6. Primary education is compulsory from ages six to twelve after which students enroll in lower secondary school. There are two types of curriculum offered in these schools, one called common general and the other vocational. An upper secondary school is available from ages 15 to 18 with halftime attendance required from ages 16 to 18. There are several different types of organizational authority for schools. They include private, usually catholic, schools, provincial schools, state schools and communal schools. The sample of students upon which the analyses for this volume is based comes from the lower secondary school.

British Columbia

For both Canadian provinces the structure, financing and control of the schools are independent of the national government. Children in British Columbia have opportunities to attend pre-schools and kindergartens prior to age 6 at which time there is compulsory enrollment in a 6 year primary school. For ages 12 to 18 there are lower secondary and upper secondary schools. Differentiation of curriculum occurs during the upper secondary schools and all children are exposed to common activities prior to then.

The sample of students comes from grade 8 which is located in the lower secondary school.

France

Schooling in France is considered to be highly centralized. Children of ages 2 to 5 may attend pre-primary schools and 1980 estimates are that about 90% of them do. Primary school extends from ages 6 to 10 and grade repetition, though declining somewhat, remains relatively common compared with other countries. The first cycle of secondary schooling is for ages 11 to 14 and contains a common curriculum except that after two years students may pursue more vocationally oriented courses. The second cycle of secondary schooling leads either to a baccalaureate degree and preparation for university or a vocational certificate. About 25% of the Population A students attend private schools. Students from the first cycle are in this sample.

Japan

Japanese children attend pre-schools and kindergartens from ages 3 to 6, elementary school from 6 to 12, lower secondary from 12 to 15 and upper secondary from 15 to 18. Upper secondary schools provide a variety of alternatives including vocational, university preparation and correspondence courses. Examinations after lower secondary school determine what upper secondary school a child attends. The sample of children in this study comes from the 1st year of lower secondary school and is comparable to grade 7 in the United States

New Zealand

Children may attend pre-school or local play centers from ages 3 to 5. Primary schooling begins on the child's 6th birthday and continues for 8 years. There are up to 5 years of secondary schooling available but students may leave earlier to pursue vocations. About 10% of the schools are private; about 30% of the schools either for boys or for girls. Students from Form 3, a kind of intermediate level between primary and secondary schools are the sampled population.

Ontario

Children may attend public supported schools as early as age 4. Primary schooling of about 6 years is followed by 6 years of secondary schooling which, in addition, contains uniquely a capstone grade 13. Ontario has both private schools and schools where the language of instruction is French. The sampled population came from grade 8, a part of the secondary school.

Thailand

Local communities provide what pre-schooling is available for students prior to the age of 6. Primary schooling extends from age 6 to age 12; lower secondary from age 12 to age 15; and, upper secondary from ages 15 to 18. Schools are financed by the national government and as indicated earlier there has been a dramatic increase in enrollments at all levels and especially at the secondary level. The sample of students comes from the lower secondary schools.

United States

A variety of pre-schooling opportunities, both private and public are available for children under the age of 6. In most states publicly funded kindergartens are available at age 5. From age 6 to about age 14 students attend elementary schools but the particular structure depends on the local school system. Secondary education through comprehensive schools that provide either college preparatory or vocation curricula are available for students of age 14 to 18. The sample of students for the study comes from the eighth grade. About 10% of the students are enrolled in private schools.

Differences in school organization may simply represent various ways to slice the same loaf of bread. That is, there may be few important consequences of having an eighth grade in a junior high versus an elementary of middle school. Yet, it is interesting to note that in the majority of countries the sample of students comes from what is perceived to be secondary schooling as opposed to the United States where students perceptually have not yet entered secondary schools.

Control of Schools

As ambiguous as school organization but more important may be the issue of control of schooling. Here, at least superficially, differences are only between the United States and the others. In the U.S. there remains, at the rhetorical level, the notion that control of schools resides in the local community. While this may be correct historically, there is little doubt that recent educational reforms have had an intended or unintended effect of diminishing local autonomy and placing more control at the state level. This change in locus comes on top of earlier efforts of the Federal Government to institute programs where in effect schools had to play by federal rules in order to qualify for federal monies. So a question for the U.S. is whether or not there is local control of schools.

Other countries presumably have central control of schools. For the two Canadian provinces central control case means provincial control. For the remainder of the countries it means that educational policies are made at a national level. Just as one can question the validity of the notion of local control for the U.S. so too it is possible to wonder what central control means for the other countries. Questions here reside around the notion of what really can be controlled. For example, no system has the power to control completely what teachers do in classrooms, but can to a greater extent control who becomes a teacher. Likewise, it is possible to define at a national level the nature of a curriculum, but it is impossible to insure that it is implemented consistently across classrooms. As will be seen later in this volume, teachers within all countries vary greatly in regard to what part of the curriculum they say they teach. In addition, most countries in the survey make provisions for local or provincial initiatives on virtually all matter related to schools. So, what one has generally is tension among various administrative levels regardless of where formal or legal control resides.

Issues of types of control over schools and how much control is possible, though complex, are crucial. Within those realms are potentially important explanatory variables. Countries use inspectors, competence examinations, school leaving examinations, financial threats, legal authority, teachers' unions and a multitude of other means to influence outcomes of schooling. Yet, prudence dictates that with limited evidence one should mention differences but not attempt to resolve them. Hence, for this volume they remain important but unresolved issues and the relationships between them and outcomes of schooling are ı .. unexamined.

Characteristics of the Studies

Countries which decide to participate in IEA studies decide what parts of the larger study they will implement. For the mathematics study additional choices were allowed in terms of how the agreed upon cognitive items would be administered. In addition, for funding and other administrative reasons some countries completed their study a year earlier than others. For these and numerous other reasons there are variations in the studies conducted by these eight countries.

The Samples

The formal definition of the students in Population A was: All students in the grade (year level) where the majority has attained the age of 13.00 to 13.11 years by the middle of the school year. As indicated earlier, this population fell in different levels of the school system depending on the structure of the schools within a country. The comparability of the samples, therefore, resides in the age of those who were sampled.

A second kind of sampling, item sampling, was conducted in the studies. Essentially, a sampled student within a country was administered a Core test of 40 items and one of four rotated forms of 35 items. So although any one student might take no more than 75 items, responses within a country would be obtained from a full set of 180 items. The pattern of items within the core and rotated forms as well as the decision of which items to administer during the pretest were left to the countries.

Both kinds of information, that of the sampling of students and the organization of the cognitive test, is given below for each of the countries.

Belgium Flemish

1. The sample: All students in the second year of the general secondary education, technical secondary education and vocational secondary education programs in both Type I and Type II forms of school organization. Less than 1% of the student cohort was excluded by this definition. The sample was composed of 168 schools, 175 classrooms and 4519 students.

2. Cognitive test: The longitudinal core was adjusted to the Belgium Flemish curriculum. Both Core and rotated forms were administered at the pretest and the posttest with complete rotation between the two occasions. Some linkage between student scores at pre and posttest times have been lost.

British Columbia

1. The sample: All students enrolled in regular grade 8 classes in September 1980 in the British Columbia public school system. Both slower students in remedial classes and students attending private schools, about 10%, of the cohort was excluded from the sample. The sample was composed of 90 schools, 93 classrooms and 2567 students.

2. Cognitive test: A standard (i.e., the same as 5 of the other countries) longitudinal core administered both at pretest and posttest. The rotated forms were given only at the posttest.

Ontario

1. The sample: Students enrolled in normal grade 8 classrooms in Ontario. The excluded population was less than 2%. The sample included 130 schools, 197 classrooms (two classrooms per school where possible) and 6284 students.

2. Cognitive test: The standard longitudinal core and rotated forms tests were administered both at pre and posttest. There was complete rotation of the forms between pre and posttest.

France

1. The sample: All students in class de 4e (grade 8) of colleges, private and public education in metropolitan France. The excluded population is estimated to be

less than 1%. The sample was composed of 184 schools, 365 classrooms (2 per school) and 8778 students.

2. Cognitive test: The longitudinal core is adjusted to the French curriculum. It was administered along with the rotated forms both at pre and posttest. Students took the same rotated form on both occasions.

Japan

1. The sample: Students in grade 1 of lower secondary school (U.S. grade 7 equivalent). Excluded were students in private schools or schools for the handicapped. About 3% of the cohort attends private schools and about 1% schools for the handicapped. The sample was 210 schools, 211 classrooms and 7785 students.

2. Cognitive test: A distinct item set. There was a special 60 item test at the pretest and then Core and rotated forms at the posttest.

New Zealand

1. The sample: All students who are in normal classes in Form 3. The excluded population was less than 1%. The sample was of 100 schools, 196 classrooms (2 per school) and 5978 students.

2. Cognitive test: The standard longitudinal Core and rotated forms were administered both at pre and posttest with complete rotation between the two occasions.

Thailand

1. The sample: All students in normal classes in grade 8 in all 71 provinces. There was no excluded population although only 85% of the cohort attends school at this level. The sample was 99 schools, 99 classrooms and 4030 students.

2. Cognitive Test: The standard longitudinal Core and rotated forms were administered both at pre and posttest. Students received different rotated forms on the two occasions with no repetition of forms.

United States

1. The sample: All students in the eighth grade of mainstream public and non-public schools. Excluded were students with disabilities sufficiently severe to require

special education classes. The sample was composed of 161 schools, 302 classrooms (2 per school) and 8372 students.

2. Cognitive Test: The standard longitudinal core and rotated forms were administered both at pre and posttests with complete rotation between the two occasions.

While these are major differences between countries in terms of sampling and types of testing, other variables also may be excluded from one or more countries. Student background questionnaires, teacher questionnaires, students perceptions of the fit of the test, use of calculators and other measures are, for the most part, present across the countries but some contain variations. Where that occurs it will be documented in the text.

The Curriculum

An IEA volume has been devoted to an investigation of the curriculum of all countries in the mathematics survey. Here a couple of instances will be used to highlight some major differences between countries. For this purpose, there are two big questions that one can ask about a country's curriculum: What is in it and which students get it?

As a partial answer to what is in the curriculum, national committees in the various countries were asked to rate items on the achievement test to determine whether or not they were appropriate. Those ratings provide a way to show how varied the curricula are in these eight countries. Table 3 contains the text for eight selected items and countries' responses to those items.

Patterns of responses across items and countries suggest that there are very different curricula despite the fact that the study deals with mathematics, a content area where it is assumed that there is so much in common. The square root item, 011, is inappropriate in both Japan and New Zealand but at least acceptable in the other countries. The two geometry items, 022 and 096, form an interesting contrast since they tend to be linked, either both acceptable or not acceptable, in Belgium Flemish, Ontario and Thailand. For the other five countries the curricula apparently includes materials related to one of the items but not the other. Item 26 could be considered either a geometry item (similar triangles) or a ratio and proportion item. Yet, it is not acceptable in either Belgium Flemish or France but fine elsewhere. The reasoning item, 114, is taught in Japan, and New Zealand as well as in Belgium Flemish and France but not in the other four countries. The item which is most generally acceptable is a probability item. Interestingly enough probability and statistics as a content area is the least

Selected Items

Item Text

011 What is the square root of 12 * 75

022 AB, CD, AD, EF are intersecting straight lines as shown

026 On level ground, a boy 5 units tall casts a shadow 3 units

044 There are 35 students in a class. 1/5 of them come to school..

055 For the table shown, a formula that could relate M and N is

096 The triangle ABC and Triangle A'B'C' are congruent and their...

114 The first error, if any, in this reasoning occurs in...

188 The picture shows some black and some white marbles. Of all...

(Add pictures of itemsl)

Rating

	BFL	BC	FRA	JPN	NZ	ONT	THA	USA
011	1	2	1	0	0	2	1	1
022	0	0	0	1	1	2	0	1
026	0	2	0	1	1	2	2	1
044	2	1	1	2	1	2	2	2
055	1	2	1	1	2	0	0	1
096	0	1	2	0	1	2	0	0
14	2	0	1	1	1	0	0	0
188	2	2	2	2	2	2	2	2

2 = Highly Appropriate 1 = Acceptable 0 = Inappropriate

Table 3. Individual countries' appropriateness ratings for selected items on the cognitive test.

represented in both the curricula of the countries and the cognitive test. Apparently, even though countries do not emphasize statistics, they do agree on what little of it should be taught.

The answer to the second question, which students are exposed to what curriculum, is straight-forward on the surface of things. For the majority of the countries, there is only one "official" curriculum for all students so the answer is that almost everyone gets the same things. Officially, there are exceptions in Belgium Flemish and the United States. Unofficially, as will be seen later in the volume there may be other exceptions. For Belgium Flemish there are two types of mathematics classes, one which is taught in the general curriculum and one that is taught in the vocational curriculum. Often these are in separate schools. In the United States, there is no official national curriculum, usually no official state curriculum but almost always different types of courses for different students within local school districts. Generally the courses are: 1) Remedial, 2) General, 3) Pre-algebra and 4) Algebra. Students typically are tracked into the various courses according to perceptions of prior achievement. This differentiation of the curriculum in the United States is the basis for a chapter later in the volume.

The International Curriculum Analysis volume gives detailed descriptions of the curricula of each of these countries. The content of the curriculum, how it is delivered and who gets it so influences what students have an opportunity to learn and do learn that it would be difficult to underestimate how important they are as explanatory variables for differences in achievements.

Conclusion

This paper is meant to provide general information about the countries. The aim is to remind the reader that there are major differences between the countries in very important ways. The analysis and interpretation of survey data is, therefore, a matter of taking things out of broader and richer contexts. With survey methodology there is no alternative to such a strategy. There is merely the necessity of reminding ourselves from whence the data came.

Footnotes

[1]Sales, Murray. New York Review of Books. April 23, 1985.

[2]Encyclopedia Britanica. 18th Edition. Chicago, Illinois.

[3]The International Mathematics Curriculum. Second IEA Mathematics Study.

[4]Husen, T. & Neville Posithewaite. International Encyclopedia of Education.

COMPLETED, ONGOING, AND PROJECTED SIMS DOCTORAL THESES

Leigh Burstein
University of California, Los Angeles

Historically, the cross-national studies conducted by the International Association for the Evaluation of Educational Achievement (IEA) have generated data bases that have fostered considerable interest over and beyond their use in the primary study reports. The broad range of questions that have been included in IEA survey and test instruments, the large, multilevel school-based samples, and the mixture of participating systems attract researchers and policy analysts with a wide array of interests. The various compendiums of IEA-linked bibliographies are testament to the fact that virtually any relevant topic which did not appear in the original study reports becomes the subject of secondary investigations by someone somewhere.

While unique among IEA studies in many respects, the Second International Mathematics Study (SIMS) is clearly attracting the secondary analysis interest of traditional IEA enthusiasts. Moreover, there have been inroads to new constituencies -- mathematics educators, analysts interested in indicator development, state educational officials, etc. If nothing else, this conference is a clear testament to the breadth of continuing interest in SIMS and SIMS related research.

My role goes beyond extolling SIMS virtues and characterizing the world of secondary analysis according to SIMS. My intent is to describe SIMS as a data resource for doctoral dissertations. At institutions where members of the SIMS curriculum and technical panels reside, there have already been a number of doctoral theses completed and others are in progress. The topics represented span a range of both methodological and substantive issues and naturally gravitate around the interests of the sponsoring professors. The largest concentrations thus far come from the Mathematics Education program at the University of Illinois at Urbana-Champaign and from the Research Methodology program at UCLA where both Bengt Muthén's and my students use SIMS to try out the latest methodological developments within a substantively rich educational database or to explore within SIMS substantive questions that have historically been investigated in the educational (school, teacher, classroom, instruction, curriculum) effects literature.

Before turning to a description of the focus and results of the doctoral dissertations, I want to condition my comments by providing some perspective on my view of their nature and purposes. Empirically grounded doctoral dissertations tend to be "constrained" investigations. Under the best of circumstances, they arise through an evolving commitment of the student to a sustained, focused line of inquiry that serves as a foundation and source of ideas (and publications) for the first phase of the post-degree career. Often, despite explicit and implicit time and resource limits, the dissertation makes an original contribution by clarifying, elaborating, or extending current thinking, or on rare occasion, by challenging conventional wisdom. The benefits to the field are both primary (knowledge production) and secondary (the development of a potentially productive new professional).

Other purposes are served as well. Quite often, the research enables the dissertation sponsor to "extend" a line of inquiry they have started and contributes to the total mosaic of the senior scholar's research domain. In most of these cases, the general idea for the dissertation topic originates from the sponsor with the student both refining the idea to reflect his or her own notions and executing the investigation under mutually agreed upon guidelines. Hopefully, the student develops enough investment in the substance of the dissertation to make it his or her own. Otherwise, the dissertation serves mainly an exercise or demonstration and thus primarily a rite of passage rather than the substantive foundation for a career.

One other general feature of the SIMS dissertations is worth noting. By necessity, these dissertations are all secondary data analysis projects. As such, the empirical investigation itself is constrained by the available quantity and quality of data. And, even in such a massive data gathering activity like SIMS, certain measures weren't included and study samples were oriented in certain ways. For example, virtually all variables in SIMS were mathematics related; even student background measures of home support and resources were linked to mathematics rather than to general encouragement and support for education. Also, even though the SIMS battery of test items was considerably larger than in previous IEA studies, for certain types of studies, item sampling from certain topics is rather sparse. Finally, as with most other IEA studies, the SIMS -data are all of the survey self-report- type. While a student- using a given set of SIMS questions and items can be expected to examine their measurement properties, they typically have little recourse if specific measures didn't work as intended.[1]

Overview of the Dissertations

A list of all SIMS dissertations completed or in progress as of January 1989 is contained in Table 1. The names of the dissertation advisor(s) are included in addition to the name of the author, year of completion, title, and institution. Where available, abstracts from the dissertations are appended.

Topically, the dissertations break down into several that are primarily methodological (Delandshere, Kao, Lehman, Ryan) and the remainder which focus on substantive aspects in mathematics teaching and learning and its measurement. Virtually every dissertation thus far has examined the achievement data in some fashion and included OTL or content coverage information. Several of the dissertations took advantage of special or unique features available in the longitudinal version of SIMS, such as the pretest data (Chang, Charles, Delandshere, Dhompongsa, Fagnano, Garnier, Hafner, Kanjanawassee, Kao) and the detailed classroom process questionnaires (Chang, Charles, Dhompongsa, Fagnano, Garnier, Hafner, Kao, Williams). Most of the dissertations used data from a single country (usually the U.S. although Thailand's data have been analyzed by three different tudents). Obviously, plenty of opportunity remains to take full advantage of the cross-national aspects of the data.

---

[1]In many instances, there were built-in redundancies in measuring certain aspects of classroom and curriculum practices that help matters somewhat. However, much of what was tried with the classroom process instruments was quite novel and thus experimental, especially for such a large study. This put students in the position of having to carry out their own validity investigations with very little literature to guide them.

Table 1. List of SIMS-related)ertations completed or in progress.

University of British Columbia

Michael K. Dirks (1986). Opportunity to learn in grade 8 schools in British Columbia. Unpublished doctoral dissertation, University of British Columbia (David F. Robitaille, Advisor)

University of California, Los Angeles

Ginette Delandshere (1986) Structural equation modeling applied to multi-level data: The effect of teaching practices on eighth grade mathematics achievement. Unpublished doctoral dissertation, University of California, Los Angeles (Leigh Burstein, Advisor)

Cheryl L. Fagnano (1988). An investigation into the effects of teachers' subject matter and subject specific pedagogy training on the mathematics achievement of eighth-grade mathematics students. Unpublished doctoral dissertation, University of California, Los Angeles (Lewis H. Solmon and Leigh Burstein, Advisors)

Helen E. Garnier (1988). Curriculum comparisons: Examination of eighth-grade mathematics instruction data from the Second International Mathematics Study in the United States. Unpublished doctoral dissertation, University of California, Los Angeles (Marvin C. Alkin and Leigh Burstein, Advisors)

Anne L. Hafner (in progress). The use of teaching method scales in exploring the relationship between mathematics teaching styles and differential class achievement. Dissertation in progress, University of California, Los Angeles (Leigh Burstein and Richard J. Shavelson, Advisors)

Sirichai Kanjanawassee (1989). Alternative strategies for policy analysis: An assessment of school effects on students' cognitive and effective outcomes in lower secondary schools in Thailand. Unpublished doctoral dissertation,

Table 1. List of SIMS-related dissertations completed or in progress. (Continued)

University of California Los Angeles (Marvin C. Alkin and Leigh Burstein, Advisors)

Chi-Fen Kao (in progress). An investigation of instructional sensitivity in mathematics achieve test items for U.S. eighth grade students. Dissertation in progress, University of California, Los Angeles (Bengt 0. Muthen, Advisor)

James D. Lehman (1986). Opportunity to learn and differential item functioning. Unpublished doctoral dissertation, University of California, Los Angeles (Leigh Burstein and Bengt 0. Muthen, Advisors)

University of Illinois, Urbana-Champaign

Chang, Li-Chu (1984). The effects of teacher and student perceptions of opportunity to learn on achievement in beginning algebra in five countries. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (James Hirstein, Advisor)

Charles, Josephine (1985). Teaching mathematics in lower secondary schools in Swaziland. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (James Hirstein, Advisor)

Dhompongsa, Gullayah (1985). The teaching and learning of mathematics in eighth grade classes in Thailand. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Kenneth J. Travers, Advisor)

Katherine E. Ryan (1987). A conceptual framework for investigating test item performance with the Mantel-Haenszel procedure. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Robert L. Linn, Advisor)

Staples, Peter M. (in progress). A study of changes in secondary school mathematics amongst nine countries between 1963 and 1983. Dissertation in progress, University of Illinois at Urbana-Champaign (Kenneth J. Travers, Advisor)

Wattanawaha, Nongsuch (1987). A study of equity in mathematics teaching and learning in lower secondary schools in Thailand. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Kenneth J. Travers, Advisor)

John B. Williams (1988). The teaching of calculus in high schools in the United States. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Kenneth J. Travers, Advisor)

Findings From Dissertations

What kinds of conclusions can be drawn from the results of the dissertations done thus far? Basically, my reading of the picture is In many instances, there were built-in redundancies in measuring certain aspects of classroom and curriculum practices that help matters somewhat. However, much of what was tried with the classroom process instruments was quite novel and thus experimental, especially for such a large study. This put students in the position of having to carry out their own validity investigations with very little literature to guide them. That substantive results tend to reinforce and elaborate points hinted at in the main SIMS volumes and in The Underachieving Curriculum is helpful. There are also consistencies with the prevailing notions in the educational effects literature on classrooms and schools and on the power of curricular opportunities (both exposure and emphasis) as a component in mathematics achievement. Examples of results along the above lines include:

1. Centrality of Content --Whether measured in terms of OTL (reported by teachers or by students), time allocations, or content emphases, the effects of the content actually covered on mathematics achievement are potent (e.g., Chang, Delandshere, Garnier, Kao, Lehman). Only prior performance (represented by the pretest) has a consistently stronger relationship to achievement (as represented by the posttest) than the content coverage measures. Even after controlling for prior performance (and thus its effects on content coverage), content coverage remains influential.

2. Influence of Textbooks -- Different types of analyses applied to data from three systems (British Columbia, Swaziland. U.S.) highlight the role played by textbooks in determining teachers' content decisions and instructional strategies, and their consequences for students (Charles, Dirks, Garnier). According to Garnier, four frequently used American textbooks at Grade 8 differed in terms of content coverage and

presentation. Students in regular classrooms using one of these texts had higher achievement scores across all content areas and performed considerably better in geometry and on items tapping comprehension and application skills. (Of course the teachers using this text tended to be older, better educated, and more experienced. These same teachers, more than other teachers, tended to emphasize problem solving skills and developing an attitude toward inquiry; they also provided more opportunity to learn and emphasized more teaching methods in all mathematics topics.)

3 Weak Effects of Teacher Training -- For the U.S. sample, at least, teacher training as typically measured by number of courses in general education, mathematics, and mathematics pedagogy has little if any impact on student learning (Fagnano). Teacher training exhibited similar associations with both pretest and posttest performance, making it difficult to disentangle the unique influence of training on student learning. There were indications that the prevalence of certain classroom processes and teacher subject matter beliefs were influenced by training in mathematics and mathematics pedagogy, but there were inconsistent results regarding the indirect effects of training (through its effects on processes and beliefs) on achievement.

4. Pervasive Influence of Prior Performance -- Regardless of the focus of the dissertation, the consequences of including the pretest to measure prior knowledge and mathematics ability were considerable. The pretest is the strongest predictor of posttest everywhere and in all content areas. It is also typically more strongly associated with most student background variables than the posttest. Consequently, controlling for prior knowledge in analyses of achievement typically eliminates the influence of most student background variables. Prior knowledge as measured by the pretest is also associated with teacher attributes, curricular opportunities, and instructional practices and processes. As a result, the effects of the latter types of variables on posttest are often dampened rather heightened by controlling for prior performance. Taken as a group, the results from the various dissertations clearly highlight the delicate task of exploring the distinction between knowing and learning (or, alternatively, status and growth) and the effects of student, teacher, class, and school characteristics on either or both.

On the methodological side of the ledger, not surprisingly, we learn that it does matter how you measure achievement and instructional experiences, and how the hierarchical, multilevel structure of the data is taken into consideration in analyses. As examples:

1. Specificity of Outcome Measures -- While some dissertations used total scores across all test items (or all items on the core) as outcomes, when subtests defined by content area or some other feature of test items are used, the patterns of relationships to

other measures (curricular opportunities, instructional practices and processes) tend to vary. Individual test items within narrowly defined content categories were differentially sensitive to student and instructional characteristics in some cases (Lehman, Kao, Hafner). Clearly, the specificity of the outcome measure mattered (as did the use of posttest only, gain, or adjusted gain).

2. Specification of Opportunity to Learn -- Once someone decided to use OTL in their investigation, the choice of which measure to use remained. Both teachers (for all items) and students (for core items at both pretest and posttest) were asked whether the content necessary to answer individual items was taught or reviewed during the year. Moreover, teachers were also asked to indicate whether the content was taught in prior years if not taught during the year. These different measures capture overlapping but nonisomorphic features of perceived content coverage. Both their interrelationships and their relationships to other variables accentuate certain aspects of the instructional opportunities experienced by students. As such their commonalities and distinctions influenced the conclusions reached in various dissertations (e.g., Chang, Fagnano, Lehman, Kao); a different choice would likely have resulted in different conclusions.

3. Choice of Relational Analysis Strategy Some dissertations conducted all relational analyses at the student level while others conducted all analyses at the class or school level. Yet others did both, or employed several variants of multilevel analysis. In studies where different analytical methods for handling the multilevel structure of the data were contrasted, the substantive interpretations suggested by different analytical strategies changed (e.g., Delandshere, Fagnano, Kanjanawassee). Typically, conducting analyses solely at the student level yielded a greater number of purportedly significant effects of class and school variables although significance levels were usually inflated in such analyses. Conducting analyses solely at the macro (class, school) level tended to mask within-school relationships of student background and prior performance measures to achievement and also interactions between student characteristics and instructional characteristics in accounting for achievement outcomes.

As reflected in the above examples, the complexity of examining survey data on teaching and learning comes through loud and clear in most of the dissertations. In most cases, the investigations started with descriptive reports of bivariate relationships among the variables of interest and proceeded to condition successively on confounding variables whose effects might have been mistakenly overlooked in interpreting a specific relationship. There are subtle intricacies in interpreting survey data and in trying to nail down such elusive constructs as teaching, instruction, curriculum, achievement, and learning. Experience can - help but may not be decisive. The dissertations examined here represent noble and often notable efforts to come to grips with complicated schooling data. That the results from specific dissertations are no more or no less definitive or illuminating than other analyses of SIMS and analyses of other data bases is to be expected. The fact remains that much has been learned from these efforts about how to address various issues using SIMS as a data resource.

Other Possible Dissertation Topics

While a lot of fertile ground has been covered by the dissertations described here, most of the issues already investigated using the SIMS data could warrant further study. Moreover, there is a considerable amount of as yet untapped territory that is conducive to dissertation research. Areas that warrant further scrutiny include the following:

1. Determinants of the Distribution of Curricular Opportunities -- I pointed out earlier that in the U.S., variables that predict posttest performance are often as highly associated with pretest performance. This pattern of results naturally leads to the Kifer's question of "who gets what?" Yet, as best as I can determine, none of the dissertations thus far and none of the SIMS analyses other than Kifer's, focuses on the factors that account for the distribution of curricular opportunities and instructional experiences. It was argued early on by Kifer and Wolfe, among others, that the most interesting relationships in the longitudinal version of SIMS would involve the pretest rather than the posttest. There are obviously competing conceptions of what constitutes appropriate mathematics for students of varying ability and prior experience levels; moreover, the prevalence of certain conceptions varies cross-nationally. In-depth consideration of competing conceptions regarding access to mathematics content and how SIMS data might illuminate them would be welcome.

2. Interconnections of Coverage and Emphasis -- For whatever reason, most of the dissertations have shied away from detailed examinations of the various ways of measuring content coverage and emphasis. (The topic-specific teacher questionnaires are still sorely underanalyzed despite the attention given them in the longitudinal volume.) What I would like to see are theory-driven conceptions of coverage and emphasis operationalized in a variety of ways and then the empirical consequences of using different operationalizations considered. So far this has been attempted mainly for OTL (e.g., Chang, Kao, Lehman).

3. "Case Studies" -- Robin's analysis in the longitudinal volume points to clusters of teachers who tend to have common beliefs and employ similar constellations of instructional practices. While certain dissertations attempted to create clusters of

teachers with similar "styles" (e.g., Delandshere, Dirks, Hafner, Williams), most have focussed on a small portion of the data provided by teachers. I have suggested before that one way of viewing the longitudinal SIMS data is as a large number of detailed "case studies". What I meant by this characterization is that each teacher provided a considerable amount of information and if these data were approached as if each teacher represented a separate case study, perhaps we could gain more insights about what constitutes the array of instructional treatments in mathematics. I could foresee identifying small subsets of teachers with specific constellations of responses to the teacher questionnaires and attempting to characterize these patterns and their consequences. Robin attempted this by brute force empirical methods but clearly more theory-driven approaches are possible.

4. Ignoring or Capitalizing on Cultural Boundaries -- In several instances, I have suggested to students that perhaps they could learn more about classroom processes by pooling SIMS data across countries. In this way, variation in instructional practices and processes increases considerably and certain contrasts can be illuminated. For example, imagine combining data from enriched and algebra classrooms in the U.S. with data from, say, French and Japanese classrooms and restricting attention to the set of test items for which most classrooms in the pooled data set had an opportunity to learn. In this data set, any culturally distinctive approaches to teaching mathematics and subsequent performance are likely to be highlighted for groups of students experiencing common curriculum intents (of course, what's excluded or not measured still matters). This is just one example of how cross-national data might benefit inquiry into issues of interest in a particular country.

5. Grade 12 -- At grade 12, there were longitudinal versions of SIMS conducted in both British Columbia and the U.S. Yet these data remain virtually unanalyzed beyond the national summary reports (Only Williams' dissertation considered grade 12 data). While there were certain structural complexities built into data collection at grade 12 that don't exist at Grade 8, regularities in instructional practices and processes are likely to be more evident.

Concluding Remarks

Again, the above do not exhaust the possible areas of fruitful dissertation investigations that could use SIMS data. Nor are these topics likely to be any easier to study than those already investigated. Nevertheless, we believe that the dissertations completed and in progress clearly attest to the value of SIMS as a dissertation resource and the value of the dissertations in achieving a better understanding of the issues that analyses of SIMS data can address. Given the array of other avenues of empirical work possible with SIMS data and that institutions beyond Illinois and UCLA might want their students to take advantage of this data resource, we have hopefully only seen the tip of the iceberg with regard to the use of SIMS data for dissertation research.

Abstracts of Dissertations

Li-Chu Chang (1984) The effects of teacher and student perceptions of opportunity to learn on achievement in beginning algebra in five countries. University of Illinois at Urbana-Champaign

The purpose of this research was to investigate the relationship between teachers and students' perceptions of the opportunity to learn and student achievement in algebra in 1 3-year-old students. Population A of the IEA Second International Mathematics Study. Two related contexts were considered: (1) student entry knowledge in mathematics and (2) the content domain being taught.

The data used in this study came from five countries: France, Japan, New Zealand, Ontario-Canada, and the United States. The data analyses were done at the class level. Four research questions were addressed in this study.

Are teacher opportunity to learn and student opportunity to learn good predictors of student achievement? Both teacher opportunity to learn and student opportunity to learn, taught this year, positively influenced achievement. However, in some countries the opportunity to learn variable had a small effect on achievement because of the homogeneity of the curriculum or the effect of having previously been taught the topic.

Which is the better predictor of student achievement, teacher opportunity to learn or student opportunity to learn? Although the opportunity to learn as perceived by teachers is consistently higher than the opportunity to learn as perceived by students in the corresponding classes, the student opportunity to learn rating is a~ better predictor of achievement gain than is the teacher opportunity to learn rating.

What is the relationship between the coverage and student achievement gain for each level of entry knowledge in mathematics? For each ability group, the mean teacher opportunity to learn score is higher than the corresponding mean student opportunity to learn score.-- Student opportunity to learn is a better predictor of student achievement gain than the corresponding teacher opportunity to learn for high and middle ability classes.

What level of coverage is optimal for student achievement gain in classes of high, middle and low knowledge in mathematics? A high student opportunity to learn rating appears to be an optimal condition for high and middle ability students. Time allocation itself was not a salient factor of achievement gain and no significant interactions between opportunity to learn and time allocation were found.

Josephine H. Charles (1985) A study of the teaching and learning of common and decimal fractions in the eighth grade in Swaziland.University of Illinois at Urbana-Champaign (James J. Hirstein, Advisor)

Purpose. The purpose of this study was to investigate the teaching and learning of common and decimal fractions in the eighth grade in Swaziland. The study focused on the three aspects of the mathematics curriculum: (a) the intended curriculum as reflected in curriculum guides, course outlines, syllabi and textbooks; (b) the implemented curriculum at the classroom level where teachers translate the intended curriculum; and (c) the attained curriculum what the students have learned as measured by the tests and questionnaires .

Procedures and Analysis. The data lend themselves to three major classifications: (a) curriculum data-context survey and textbook analysis; (b) classroom data include the Teacher, Topic and Attitude Questionnaires; and (c) student data include cognitive and attitude data.

The definition of Population A was modified for Swaziland as the grade level where 13 year-old students should be found according to the school system. A pre-test was administered to 904 students in 25 classrooms in February 1980 and a posttest in September 1980. The teachers responded to the Classroom Processes Questionnaire for common and decimal fractions. Results of the Teacher Questionnaires and student achievement tests were analyzed using Pearson's Correlation and ANOVA.

Selected Findings and conclusions. The classes were identified as remedial~ typical enriched or accelerated with an average class size of 27. An equal amount of time is spent on fractions and other topics in the mathematics curriculum. The majority of the teachers were young and inexperienced. Much of the teachers' time is spent on presenting new content or reviewing old material and a relatively small proportion of time is spent on discipline or administration tasks. The textbook provided the "boundaries" for what is taught. Limited use is made of resources beyond the textbook for either content or methods of teaching. The majority of student time is spent listening to teacher presentation, doing seat work or taking tests. Little time is spent on group work. Instruction in fractions tends to be symbolic and formal with an emphasis on computational proficiency. Students' performance is higher on common fractions and on application level items. Both the teachers' attitudes and beliefs and students attitudes and beliefs had no effect on student achievement.

Ginette Delandshere (1986) Structural equation modeling applied to multilevel data: The effect of teaching practices on eighth-grade mathematics achievement. Unpublished doctoral dissertation, University of California, Los Angeles (Leigh Burstein, Advisor)

Student achievement is mostly affected by three types of variables: student ability or aptitude, student characteristics (i.e., home background), and various combinations of teacher and classroom characteristics.

The present study questions the adequacy of the analytical models traditionally used in school and classroom effect research. It is assumed here that the variability in the relationship between student characteristics and achievement could be more effectively examined as a function of students' instructional experience. The analytical scheme proposed here is intended to reflect the multilevel nature of the data, to take measurement error into account, and to allow for the examination of the interrelationship among the predictors of student achievement.

The investigation is carried out with data collected from students and teachers in 226 U.S. eighth-grade mathematics classrooms (Second International Mathematics Study under the auspices of IEA). The analytical scheme tested here includes the following steps: 1) classification of teachers according to instructional practices using three clustering algorithms (K-means, Ward's method, and NORMIX), 2) comparison of the effect of group membership defined by clustering on achievement to more traditional methods (regression and ancova), 3) estimation of a student achievement model (using LISREL) within each group as defined by clustering, and 4) comparison of the model across groups to assess the structural differences in student achievement due to differences in instructional practices.

A five cluster solution was retained, and cluster membership was found to account for an amount of variance comparable to that which would be explained by regressing achievement directly on the teacher variables used to identify the clusters.

Structural equation modeling was then used to fit a student achievement model separately in each cluster. A good fit was obtained for the model in at least three of the clusters. Finally, a multiple group analysis was conducted on the three clusters, revealing differences in the structural parameters across groups.

Guliaya T. Dhompongsa (1984). The teaching and learning of mathematics in eighth grade classes in Thailand. University of. Illinois at Urbana-Champaign

This study surveyed and analyzed data relating to classroom processes and student achievement in mathematics in Thailand. It also inquired into the relationships between such processes and achievement, and investigated the differences in instructional behaviors among teachers whose students exhibited low learning gain. Furthermore, the study examined factors affecting student achievement in mathematics.

The study was conducted in Thailand in conjunction with the Second IEA International Mathematics Study. The sample, drawn through the use of the probability proportional to size (PPS) sampling procedure, consisted of 45 classrooms from 23 schools in 10 provinces, with two classes per schools and a total of 1,910 eighth grade students. The data collected included students' pretest and posttest achievement, classroom processes reported by the teachers, and information on student home background, teacher characteristics and school conditions. These data were obtained through the administration of relevant tests and questionnaires.

Descriptive results regarding the ways the teachers provide instruction of ratio, proportion and percent were reported both verbally and graphically. Some of the more important findings obtained from the multivariate analyses are as follows: 1) student prior knowledge in mathematics and consistency of instruction contribute the most to student post-achievement variance. 2) The variables associated with high-gain teachers seem to be consistency of instruction, use of class time in explaining new content and in managing the classroom and emphasis on practice and drill more than on problem solving. 3) The variables associated to low-gain teachers seem to be the use of a variety of teaching techniques and the emphasis on problem solving more than of practice and drill. 4) Students' prior knowledge of mathematics appears to affect students' final achievement in mathematics directly and strongly, while the classroom process factors seem to have negligible effect on achievement. Other background factors show minimal indirect effect on achievement, but home status and processes in the home strongly and directly affect student prior knowledge in mathematics, which, in turn, affect students' final achievement.

Michael K. Dirks (1986). The operational curricula of Mathematics 8 teachers in British Columbia. The University of British Columbia, Canada (David F. Robiraille, Supervisor)

The purpose of this study was to describe the mathematics curricula as actually implemented by a sample of Mathematics 8 teachers in British Columbia. A survey of previous research indicated that knowledge about the mathematics subject matter which teachers present to their students and the interpretations which teachers give to that subject matter is sparse in spite of the importance such knowledge might have for the curriculum revision process, textbook selection, the identification of in-service education needs, and the interpretation of student achievement results.

The mathematics 8 curriculum- was divided into three content areas: arithmetic, algebra, and geometry. Within these content areas a total of 16 topics were identified as among the basic topics of the formal Mathematics 8 course. Four variables were identified as representing important aspects of a mathematics curriculum. The first of these, content emphasis, was defined as a function of the amount of time a teacher spent on each content area. The other three variables, mode of content representation, rule-orientedness of instruction, and diversity of instruction, were defined as functions of the content-specific methods teachers used to interpret the topics to their students.

Class achievement level and the primary textbook were identified as having strong potential relationships with a teacher's operational curriculum. These were used as background variables in this study.

The data for this study were collected as part of the Second International Mathematics Study during the 1980-1981 school year. The sample consisted of 93 teachers who submitted five Topic Specific Questionnaires throughout the school year regarding what they taught to one of their Mathematics 8 classes.

Among the findings of this study were: (1) Wide variation existed in the emphasis given by teachers to the three content areas with 60% giving at least one area light or very light emphasis. (2) Teachers using a text which placed more emphasis on a particular content area tended to spend more time on a particular content area in their classes. (3) Teachers of low achievement classes tended to present mathematics in a slightly more abstract and rule-oriented way than teachers of high achievement classes.

Cheryl L. Fagnano (1988). An investigation into the effects of teachers' subject matter and subject specific pedagogy training on the mathematics achievement of eighth-grade mathematics students. Unpublished doctoral dissertation, University of California, Los Angeles (Lewis H. Solmon and Leigh Burstein, Advisors)

This study addresses the empirical questions; does the amount and kind of training a mathematics teacher acquires affect his or her choice of classroom processes, pedagogical beliefs, and ultimately student's mathematics achievement? The United States 8th grade sample from the Second International Mathematics Study was the data source. Using a two stage analysis, four models of multiple regression were used to investigate the study's hypothesis. Three outcome measures were investigated, student posttest scores, classroom processes, and teacher pedagogical beliefs. The major independent variables of interest were three types of teacher training, subject specific, pedagogical, and general education. The study's most significant finding was that increased amounts of pedagogical training was negatively associated with student achievement. This finding while suggestive was inconclusive due to problems of multi-colinearity between measures of teacher and
student quality.

Helen E. Garnier (1988). Curriculum comparisons: Examination of eighth-grade mathematics instruction data from the Second International Mathematics Study in the United States. Unpublished doctoral dissertation, University of California, Los Angeles (Marvin C. Alkin and Leigh Burstein, Advisors)

Previous research has identified the classroom textbook as the major contributing factor to determining what teachers teach and what students learn. Given the enormous potential of the textbook to guide instructional processes, the textbook is an essential variable to be included in any comparison of different curricula. National mathematics data from the Second International Mathematics Study provided information on students, teachers, and instructional processes. The four most frequently used U.S. eighth grade mathematics textbooks from that study were used to investigate curricula characteristics. The extent to which different textbooks influenced different instructional processes and different patterns of student achievement were examined.

Qualitative comparisons of the textbooks indicated both differences in content coverage and presentation. Descriptive analyses of student, teacher, and instructional process measures identified statistically significant differences in mathematics curricula. The degree to which student, teacher, and instructional process variables explained variation in mathematics achievement scores also differed across the curricula defined by textbook choice.

The results of the analyses provided essential information about evaluation questions of effectiveness and causality. Students in typical classrooms using the more advanced mathematics textbook had significantly higher mathematics achievement scores in arithmetic, geometry, measurement, and algebra. They had the largest gains in geometry, and in comprehension and application skills. Teachers using the more advanced textbook were the oldest, most experienced, and most educated teachers. They emphasized problem solving skills and developing an attitude of inquiry more than other teachers. They provided more opportunity to learn and emphasized more teaching methods in all mathematics subjects. These teachers used self-written materials more than other teachers.

The analyses suggest further studies might be done on the contribution of teacher and instructional process variables to explain the variation in mathematics achievement scores for remedial and enriched students. Also more detailed analyses of mathematics topics within arithmetic, geometry, measurement, and algebra are suggested.

Anne L. Hafner (in progress). The use of teaching method scales in exploring the relationship between mathematics teaching styles and differential class achievement. Dissertation in progress, University of California, Los Angeles (Leigh Burstein and Richard J. Shavelson, Advisors)

The purpose of this study is to examine the influence of specific teaching practices on class-level student mathematics achievement. Prior studies have identified general teaching behaviors which are related to student achievement, but not math-specific behaviors. The major contribution of this study is to identify teaching styles/practices in the mathematics content domain which influence class mathematics achievement. In addition, the study will attempt to disentangle the OTL (content coverage) influence from the teaching method influence.

The study tests the hypothesis that teaching practices will influence differential content coverage (OTL) above and beyond the influence of prior class achievement and background variables. It also hypothesizes that after controlling for OTL and background variables, differential performance between classes will still exist which may be attributable to teaching styles or practices. Finally it is hypothesized that classes taught by various teaching "styles" will show differential achievement, and that practices which focus on perceptual presentation, which stress an informal approach that links across mathematics concepts and which use multiple concept interpretations will best predict high achievement.

Sirichai Kanjanawassee (1989). Alternative strategies for policy analysis: An assessment of school effects on students' cognitive and effective outcomes in lower secondary schools in Thailand. Unpublished doctoral dissertation, University of California Los Angeles (Marvin C. Alkin- and Leigh Burstein, Advisors)

The purpose of this study is to consider alternative multilevel strategies to assess the school effects on various dimensions of student outcomes. The present study questions the adequacy of the conceptual analytical models used in school effectiveness research. The conceptual strategies proposed here were intended to obtain a relevant model which reflects the multilevel nature of educational data, while the analytical strategies which take into account the multilevel structure were aimed to allow for the variation of coefficient estimates between levels, and to test the fit of school effect models. The investigation was carried out with data from the Second International Mathematics Study (SIMS) collected in Thailand from 4,030 eighth grade students and their mathematics teachers and administrators in 99 schools. The analytical strategies to detect, explain, and compare the school effects included variance component analysis, standard regression analysis, hierarchical analysis of covariance, and selected multilevel analysis techniques (OLS single equation, OLS separate equation, and HLM approaches.) The major findings can be summarized as follows. 1) The alternative strategies for traditional multilevel analyses are needed in order to provide more realistic, informative, and accurate assessment of school effects. 2) Thai schools did differ in enhancing students' status and growth in cognitive and affective mathematics outcomes. 3) The outcome variables were affected by multilevel variables: student backgrounds, class/school characteristics, and socio-cultural contexts. 4) The important variables affecting the outcomes were students' prior achievement and attitude, expectation for further education, use of home calculator, parents' contribution to the learning, parents' motivation, peers' achievement, class size, teacher experience, student-teacher ratio, and qualified mathematics teacher ratio. The student-backgrounds tended to have strong effects on students' status in cognitive and affective outcomes, whereas, the class/school characteristics tended to have strong effects on students' growth in cognitive and affective outcomes.

Chi-Fen Kao (in progress). An investigation of instructional sensitivity in mathematics achieve test items for U.S. eighth grade students. Dissertation in progress, University of California, Los Angeles (Bengt O. Muthén, Advisor)

The purpose of this dissertation is to further elaborate and study the applicability of the extended IRT model developed by Muthén (1987). Muthén's approach allows for the incorporation of auxiliary information about the background and characteristics of students in the estimation of an IRT measurement mo The effects of auxiliary variables on ability estimates and the effects of ability and auxiliary variables on performance can be estimated within a common modeling framework.

The dissertation focuses on refinements in the investigation of the instructional sensitivity of test items using the SIMS data base. In earlier analyses family background and item specific opportunity-to-learn (OTL) information were used in studying performance on the items from the core test. The work is expanded in the following ways: (1) the analyses will be done with the pool of 180 items from both core and rotated forms with procedures developed to handle the "random missingness" involving the rotated forms; (2) the array of instructional variables will be extended beyond item-specific OTL; (3) new ways will be developed to handle OTL other than as item specific influence.

The study attempts to answer such questions as:

Do instructional coverage effect achievement performance in addition to its effects on latent ability?

If an item is instructionally sensitive, is it still a good measurement of the ability?

What kinds of items tend to be sensitive?

James D. Lehman (1986). Opportunity to learn and differential item functioning. Unpublished doctoral dissertation, University of California, Los Angeles (Leigh Burstein and Bengt 0. Muthén, Advisors)

This student was intended to examine the impact of differences in opportunity to learn (OTL) item content on the functioning of items and the degree to which such differences can lead to improved understanding of the results from investigations of item bias. The present study sought to demonstrate two things: (l) student differences in opportunity to learn item content cause differential item functioning (DIF); and (2) statistical indications of item bias (in the present case associated with gender) confound differences in item functioning attributable to gender with those due to differences in opportunity to learn. An Item Response Theory approach to item bias and differential item functioning was used to address the questions of the study. The data source was a sample of eighth grades in the U.S. who participated in the Second International Mathematics Study (SIMS). The items investigated were taken from the 40-item core test of mathematics. The analysis focused primarily on algebra items from this test because of the substantial variability of OTL across students in this topic area.

The primary results can be summarized as follows: (l) All eight algebra items exhibited differential item functioning associated with differences in OTL. Specifically, the item characteristic curves for high and low OTL groups indicated that students of a given ability level in the OTL groups had a higher probability of getting the algebra items correct than members of the low OTL groups. (2) Evidence of possible gender bias was found in only two of the eight items. Thus it was not possible to conclude that OTL DIF confounds gender DIF. The lack of confounding must also be attributed to the very similar levels of OTL between boys and girls. However, OTL DIF in this population on this type of test was clearly shown.

Katherine E. Ryan (1987). A conceptual framework for investigating test item performance with the Mantel-Haenszel procedure. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Robert L. Linn, Advisor)

Recently, the Mantel-Haenszel (MH) procedure has been suggested as an alternative procedure to IRT methods for investigating item bias (Holland & Thayer, 1986; McPeek & Wild, 1986). However, there are few studies examining the stability of the MH procedure across different samples of test takers (See McPeek & Wild, 1986). No studies have examined whether the Mantel Haenszel estimates are stable within different sets of items. This study examined the stability of the MH estimates across different samples of test takers as well as across different sample sizes: investigated whether the MH procedure is robust with respect to item context effects; and whether the identification of differential item functioning can be improved by controlling for the multidimensionality of the matching criteria by controlling on an additional criterion. Results indicated that a sample of 6000 for black-white comparisons was not adequate for obtaining stable estimates from the MH procedure while the MH odds ratio appears to be robust to item context effects.

Nongnuch Wattanawaha (1986). A study of equity in mathematics teaching and learning in lower secondary school in Thailand. University of Illinois at Urbana-Champaign (Kenneth Travers, Adviser)

The major purposes of this study were 1) to assess the extent to which the recently reformed (1978) Thai national curriculum has been implemented by teachers in different regions of the country, 2) to assess the extent of variation of student achievement across regions and across schools, and 3) to explore some determinants of achievement patterns that are potentially within the control of the school system, particularly content coverage and classroom practice. This study was undertaken as part of the Second International Mathematics Study (SIMS), using international data as well as national data.

The findings suggest that 1) There are no significant differences in the coverage of major areas of content across any of the analytic units investigated in this study, educational regions, social-cultural contexts, and classrooms of different achievement levels. 2) There are no significant differences in student achievement among educational regions but there are differences at the class level which tend to be associated with rural and urban environments. 3) High-achieving classes and low-achieving classes do not vary in content coverage, but do show patterns of differences which can be interpreted in terms of conceptic : of active teaching proposed by Good, Grouws, and Ebmeier (1983).

The design of SIMS permitted a comparison of Thailand with other nations. Thai national achievement is lower than that of most other nations, but when the achievement of students in Bankok is compared with other nations, the ranking is similar to that of the United States of America and New Zealand.

John B. Williams (1988). The teaching of calculus in high schools in the United States. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign (Kenneth J. Travers, Advisor)

This thesis utilized data from the Second International Mathematics Study to characterize U.S. high school calculus classes and to identify aspects of teachers and teaching of calculus that accounted for differences in class achievement. The factors examined were (a) the degree to which teachers' presentations of mathematics were process oriented, (b) the degree to which teachers used formal methods of instruction, (c) the extent to which teachers relied on the textbook, (d) the percent of time that the class spent working in small groups, and (e) the percent of time that students spent working alone. A detailed profile of high school calculus teachers and classes was developed, including such variables as teacher background, curriculum content, manner of teacher presentations, and decisions regarding the teaching of the target class.

Of the five factors and their interactions, none showed a significant relationship to achievement. Exploratory analyses suggested that classes which spent less time in small groups showed a greater achievement in comprehension accompanied with higher variance in overall achievement. The data suggested that greater teacher reliance on the textbook coupled with more time spent in small groups was associated with lower achievement at both the computation and comprehension levels. Finally, teacher presentations containing formal proofs were associated with greater variance among classes at the higher cognitive levels of achievement. Implications for future development of the high school curriculum in calculus are included.

CONTENT REPRESENTATION IN COLLEGE ALGEBRA: SUMMARY REPORT

Peter Lochiel Glidden
University of Illinois at Urbana-Champaign

Abstract

The *Second International Mathematics Study* College Algebra Classroom Process Data for Population B were examined: (a) to study reasons cited by teachers for teaching subtopics, (b) to study reasons cited for selecting particular content representations, and (c) to determine what relationships exist, if any, between teachers who use multiple content representations and their teaching decisions, professional opinions, backgrounds, classes, and schools. Important differences were found in the reasons cited for and against teaching subtopics. Considerable differences were found in teacher choice of representation and in reasons cited for and against use of a particular representation. Relationships were found between teachers who use multiple representations and: (a) their development and use of supplemental materials, (b) their presentation of content, and (c) their sources of ideas for applications. Evidence was found relating the use of multiple representation to teacher background and education.

The method or strategy used to present or interpret a mathematical concept is important for curriculum designers, textbook authors, and classroom teachers. This finding is consistent with and supported both by traditional learning theories (e.g., Ausubel, 1968; Piaget, 1975; Novak 1977) and any general or extensible cognitive science model of learning (e.g., Winston,1972; Lebowitz, 1983).

McKnight and Cooney (in press) examined content representation for Population A for all systems completing the classroom process surveys. Their study investigated various aspects of representation used including: use of symbolic vs. perceptual representation, variety of representations used, balance between symbolic and perceptual representations, and teacher opinions vs. representations used. They found some evidence of a relationship between time allocated for instruction and variety of representations used and some individually interesting results about teacher opinions. But no clear, overall patterns emerged from the data.

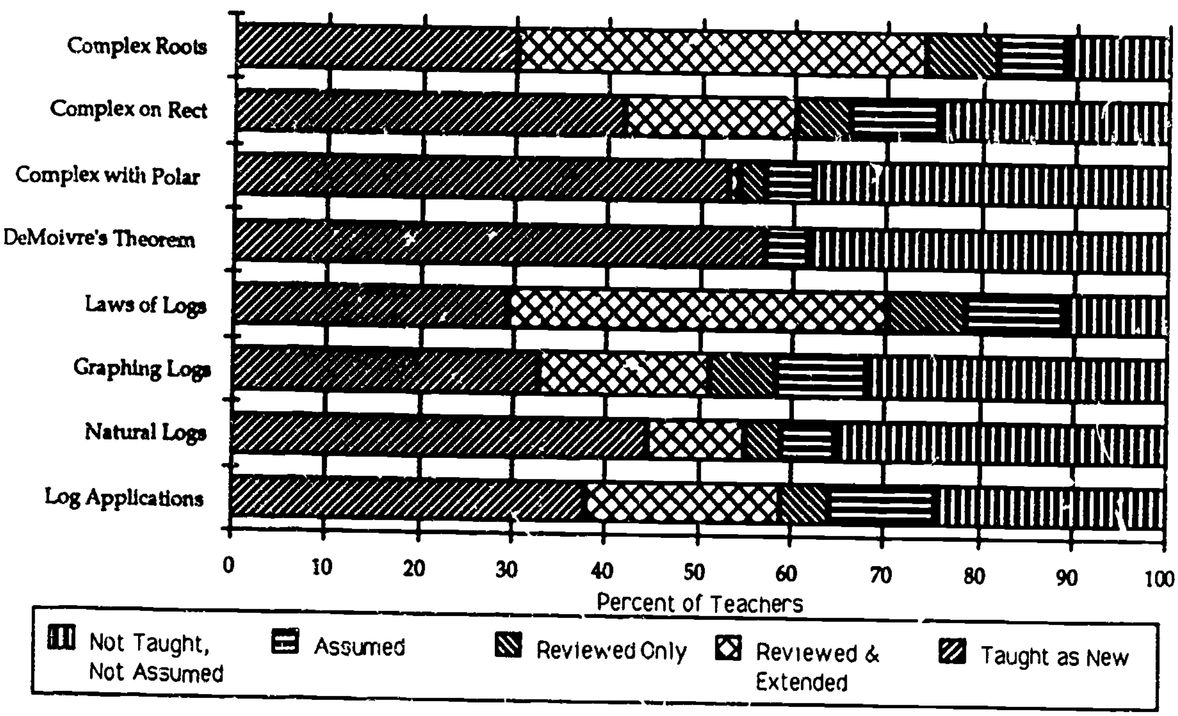
Goals of this Analysis

Instead of examining content representations across systems, we shall increase the magnification of our microscope and consider content representations for Population B College Algebra in the United States. The major goals of the present study are threefold: (a) to examine how logarithms and complex numbers are taught; (b) to determine why teachers choose particular concept representations; and (c) to determine what relationships, if any, exist between teachers' use of multiple content representations and their teaching decisions, professional opinions, backgrounds, classes, and schools.

What is Being Taught about Complex Numbers and Logarithms

Figure 1 illustrates subtopic coverage for the topics of logarithms and complex numbers. (N teachers = 153) A subtopic such as complex roots of quadratic equations is "covered" if it has been taught as new or reviewed and extended or reviewed only. A subtopic is "not covered" if it is assumed or not assumed and not covered. (Full descriptions of the labels along the vertical axis are given in Appendix 1.) Polar coordinate representations of complex numbers and DeMoivre's Theorem overwhelmingly are taught as

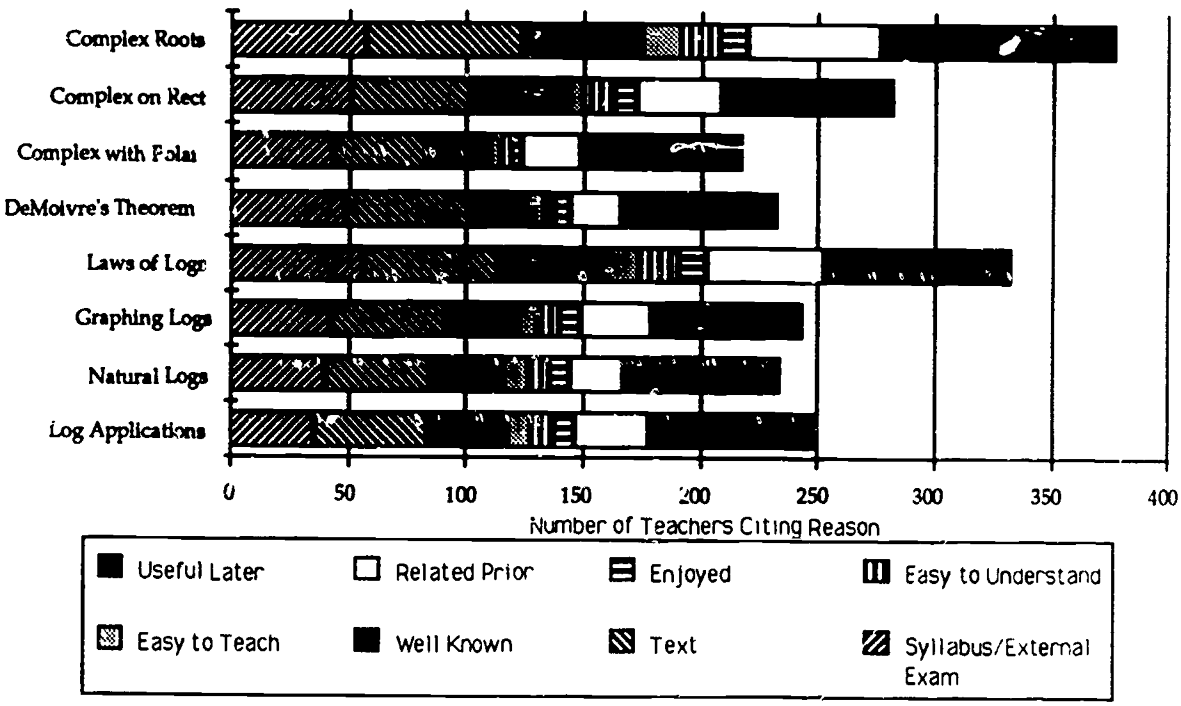
Figure 1
Coverage of Logarithm and Complex Number Subtopics



new, and complex roots of quadratic equations and laws of logarithms are almost equally taught as new and reviewed and extended. The remaining subtopics are covered mostly as new.

Figure 2 illustrates the positive reasons given by teachers for teaching the subtopics. Teachers were asked to mark as many reasons as applied. (For this and the remaining figures the subtopics are displayed in order of decreasing coverage within each topic.) The subtopics most often taught usually have the most reasons cited why the subtopic should be taught. For all subtopics, **useful later** is the most frequently cited reason followed by **text** (for six out of eight), **syllabus/external examination**, the subtopic is **well known** to the teacher, and the subtopic is **related to prior** mathematics. With the single exception of DeMoivre's Theorem, mathematical content reasons (**related to prior** or **useful later**)

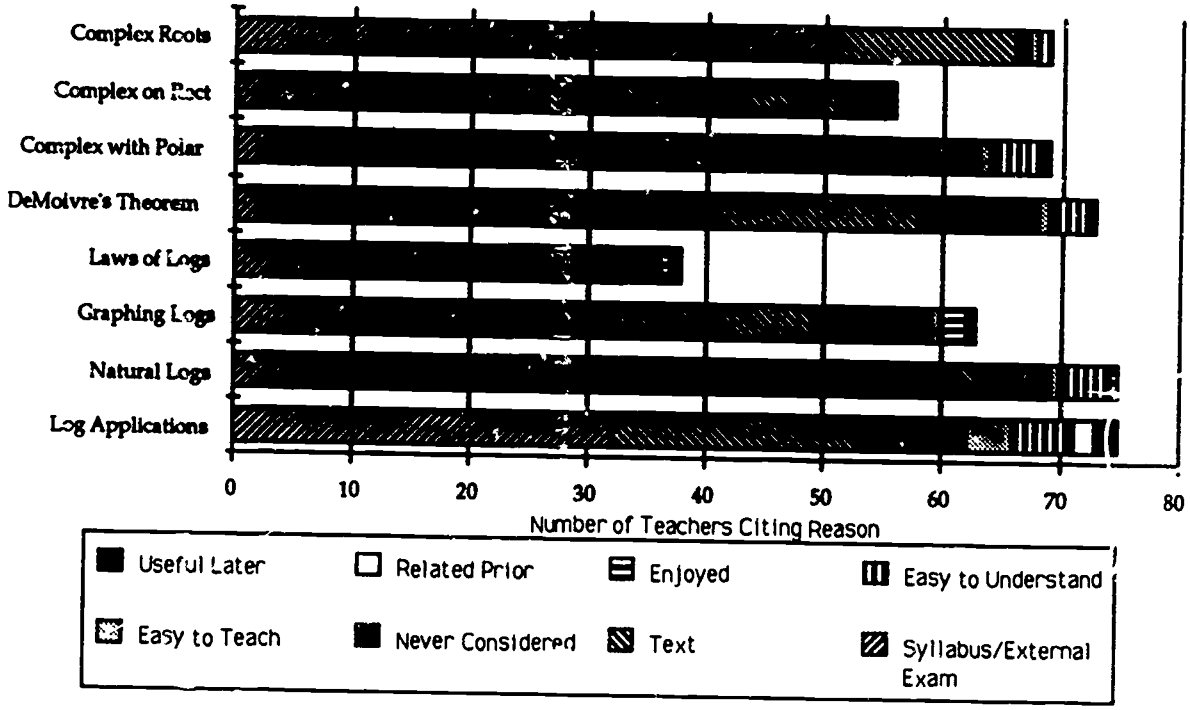
Figure 2
Positive Reasons Given for Teaching Subtopics



consistently provide most of the reasons for teaching subtopics. These reasons are followed closely by external reasons (**text** and **syllabus/external examination**).

By contrast, Figure 3 illustrates reasons cited for <u>not</u> teaching a subtopic. As might be expected, subtopics taught less frequently have more reasons cited for not teaching them than frequently taught subtopics. Overwhelmingly, teachers cite external reasons (sylla-

Figure 3
Reasons Given for NOT Teaching Subtopics



bus/external examination most frequently, followed by text) for not teaching subtopics with never considered playing a supporting role. Easy to teach, enjoyed by students, or easy for students to understand rarely are cited either for or against teaching particular subtopics. Therefore, the data in Figures 2 and 3 suggest that for a teacher to decide to teach a subtopic, not only must the subtopic be included in the syllabus or text, but the teacher must be familiar with the topic, know how the topic will be useful later, and know how the topic relates to prior mathematics.

Concept Representations

Description of Representations

Complex Numbers

The following four interpretations of complex numbers were considered in the survey (SIMS, 1985):

$i^2 = -1$. From $x^2 + 1 = 0$, we define $i^2 = -1$ and then use the distributive property to give a rationale for the product:

$(a + bi)(c + di) = ac + bdi^2 + bci + adi = (ac-bd) + (bc + ad)i$

Dilation, Rotation. Multiplication is considered as a rotation transformation followed by a dilation (stretch or shrink) transformation.

If $z_1 = a + bi = r_1 (\cos a + i \sin a)$ and $z_2 = c + di = r_2 (\cos b + i \sin b)$

then $z_1 z_2 = r_1 r_2 \{\cos(a+b) + i \sin (a+b)\}$

Dilation Rotation

Definition of Multiplication. Multiplication is defined by stating $(a + bi) (c + di) = ac-bd) + (bc + ad)i$ and then it is verified that multiplication in C satisfied the various Algebraic properties of a field.

Ordered Pair. Multiplication is defined as follows:

If $z_1 = (a,b)$ and $z_2 = (c,d)$

then $z_1 z_2 = (ac-bd, bc+ad)$

After the definition is stated, the operation is checked to see if multiplication thus defined satisfies the algebraic properties of a field.

As Figure 4 shows, $i^2 = -1$ was overwhelmingly (over 60%) the most frequently used interpretation while **dilation, rotation** was the least used (not used by 67% of the teachers). The other two interpretations, **ordered pair** and **definition of multiplication**, were used frequently or infrequently by 45% and 63% of the teachers, respectively. (For Figure 4 and the remaining figures of this section, the interpretations are ordered from left to right in order of decreasing frequent use, which coincidentally is the same order as frequent/infrequent use.) Overall, teachers used the same concept representation for all students rather than differentiating by ability.

Figure 5 illustrates positive reasons cited for using a particular concept representation. These reasons follow much the same pattern as the reasons for covering a subtopic, with one notable exception. As with subtopic coverage, the number of reasons cited correlates directly with the number of teachers who used the interpretation and the specific reasons frequently cited are content (**uses prior, useful later**) and external (**text, syllabus/external examination**) with **well known** again playing a supporting role. For concept representation, however, **easy to understand** and **easy to teach** frequently are cited, but they are not often cited as a reason for teaching a subtopic. As Figure 6 shows reasons cited for not using a particular interpretation largely were external (**text, syllabus/external examination**) and **never considered** with **prerequisites unknown** listed for **dilation, rotation**.

Figure 4
Interpretations Used for Complex Numbers

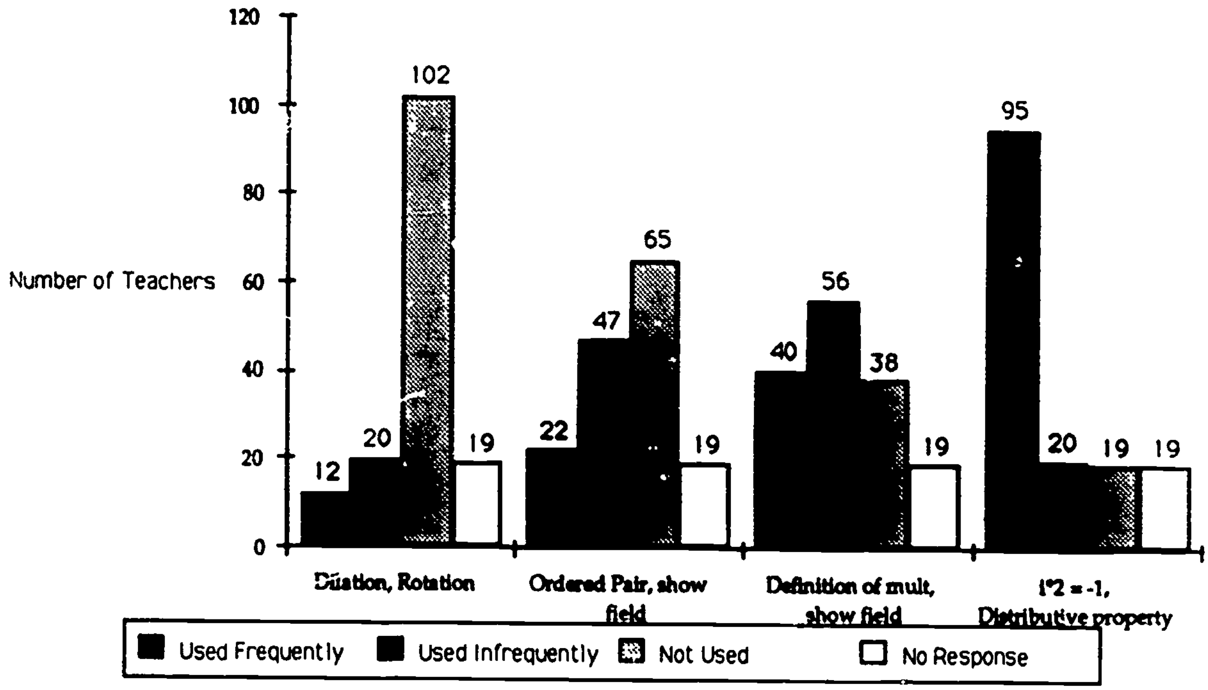


Figure 5
Positive Reasons for Using Interpretations

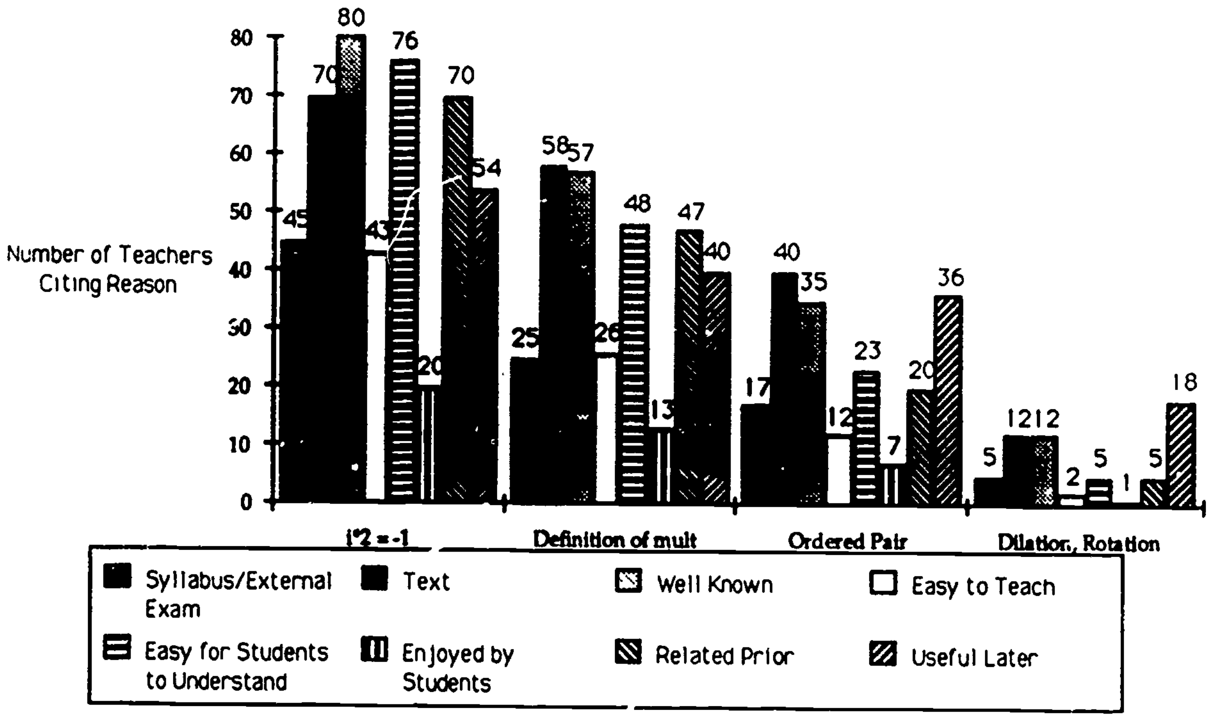
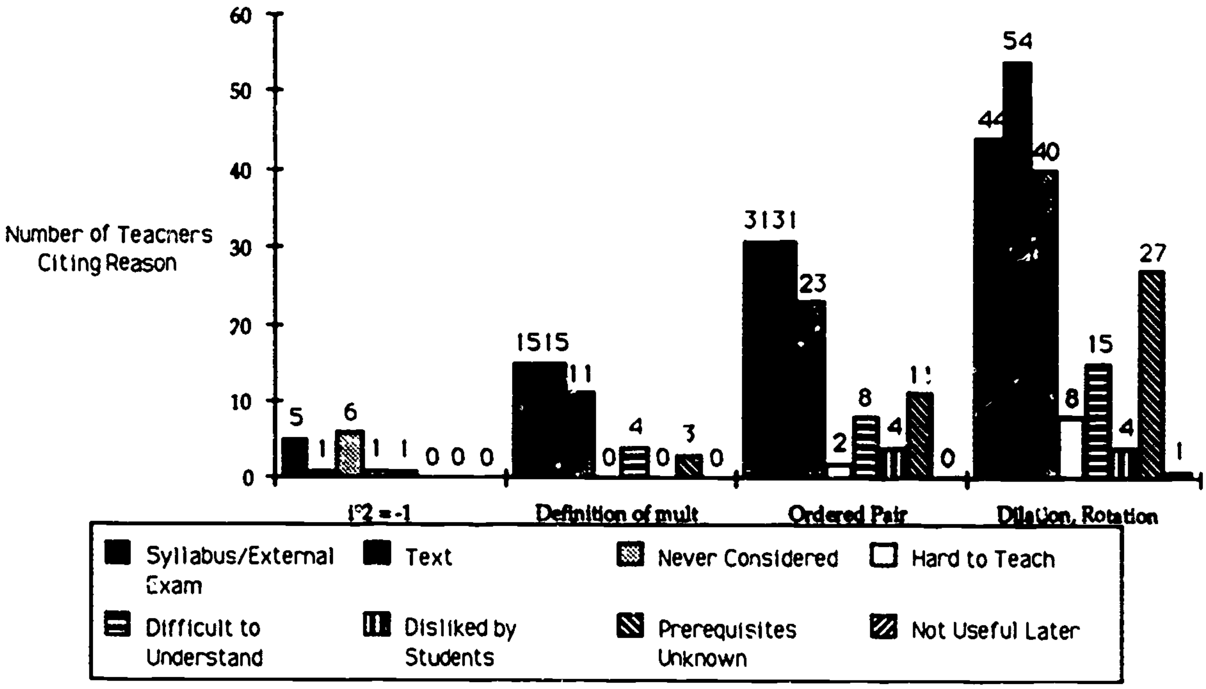


Figure 6
Reasons Why Interpretations Were NOT Used



<u>Logarithms</u>

The following four representations were considered in the survey (SIMS, 1985):

Exponent Base. Logarithms are defined as exponents., Students abstract the generalization from observing, and working with, patterns such as

$4 \times 32 = 2^2 \times 2^5 = 2^7 = 128$

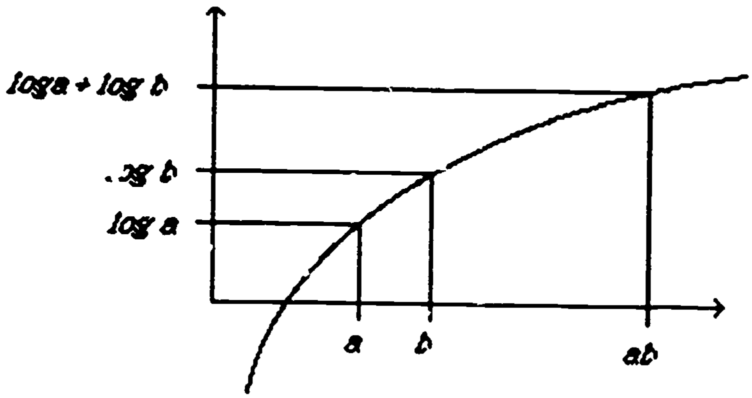
here $\log ab = \log a + \log b$ is considered a restatement of $10^a \times 10^b = 10^{a+b}$.

Inverse Function Base. A logarithmic function is defined as the inverse of the exponential function

$f(x) = 10^x$

Consider the graph of the log function. It is observed for several specific problems that the ordinate at $x = ab$ is equal to the sum of the ordinates at $x = a$ and at $x = b$. Thus

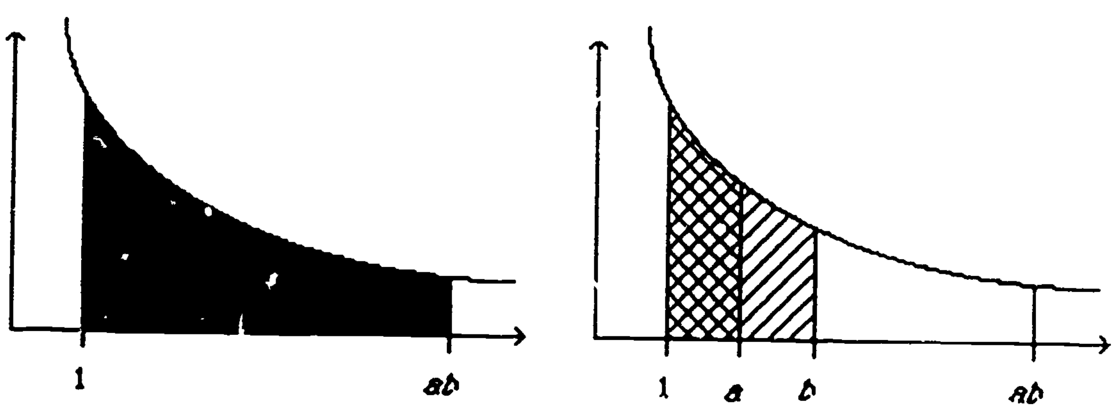
$\log ab = \log a + \log b.$



Area Under a Curve Base. Logarithmic functions are defined in terms of area under curves of the form

$f(x) = F(k,x)$ $\quad$ ($y = \log x$ is associated with $k = 0.434$)

Log b is then defined as the area under the graph of $f(x)$ for $1 <= x <= b$. By counting squares on a fine grid paper for several problems, students for [sic] the generalization that the area under the curve from 1 to ab is the sum of the area under the curve from 1 to a and from 1 to b.



As with complex number representations, teachers by and large did not differentiate representations by student ability, and one particular representation dominated (exponent base) and one representation rarely was used (area under curve) (See Figure 7). (For Figure 7 and the remaining figures of this section, the representations are displayed in order of decreased use.)

Figure 7
Time Spent on Each Logarithm Representation

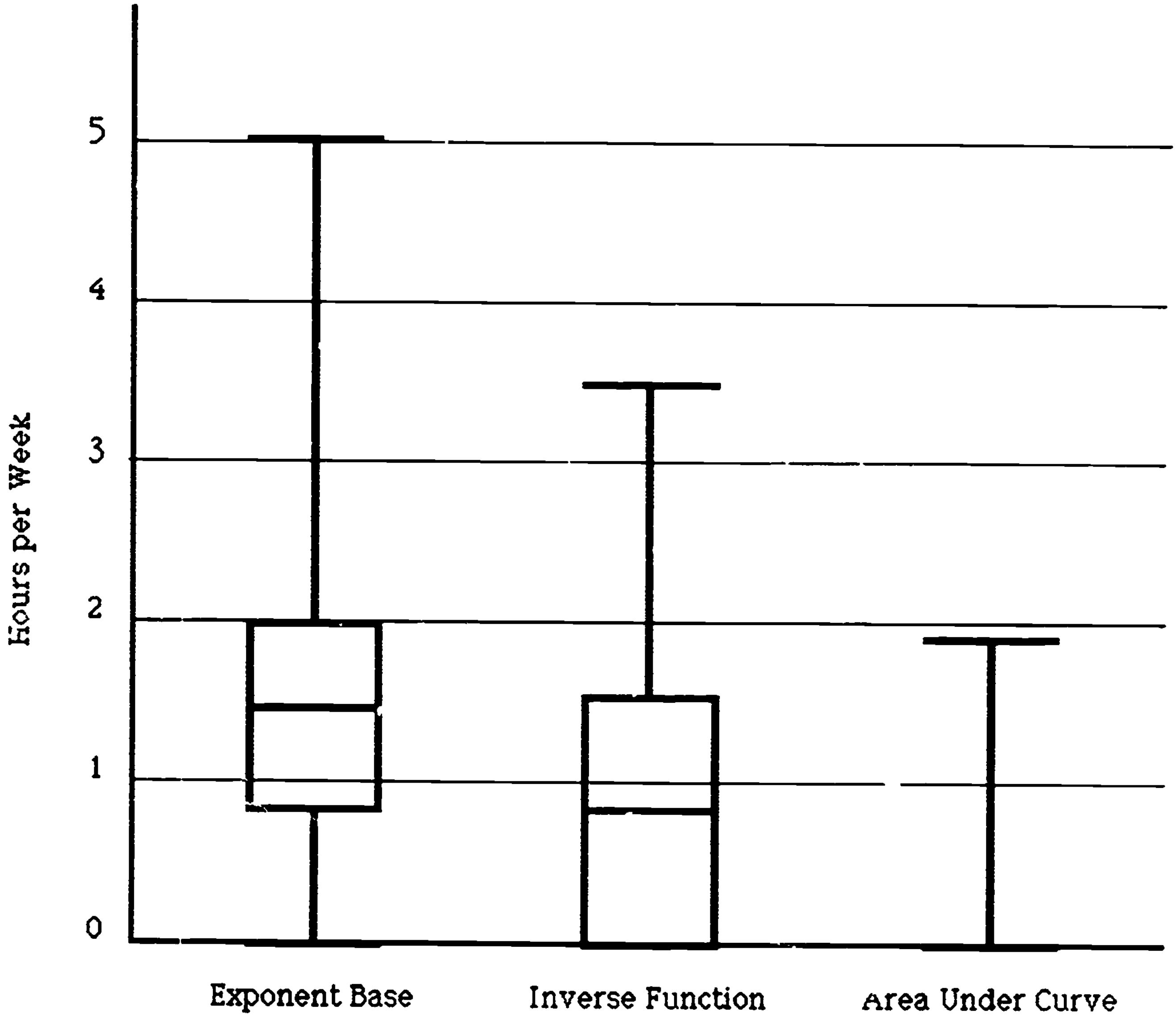
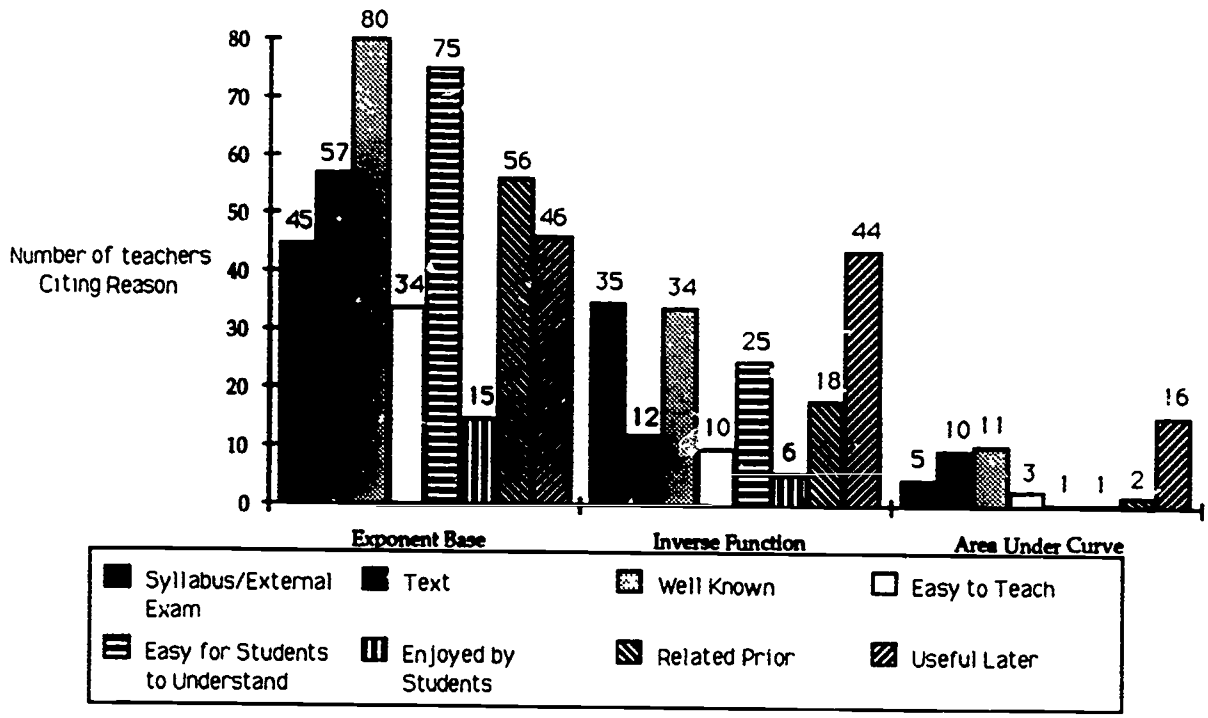


Figure 8 illustrates the number of positive reasons cited for using each representation. As seen before, there is a direct relationship between reasons cited and representation use. Paralleling complex representations, **enjoyed by students** and **easy to teach** were cited rarely and external (**text, syllabus/external examination**), content reasons (**related prior, useful later**), **well known**, and **easy to understand** were cited frequently.

The reasons often cited for not using a logarithm representation also closely parallel negative reasons for complex number representations. (See Figure 9.) The negative reasons

Figure 8
Positive Reasons for Using Representation



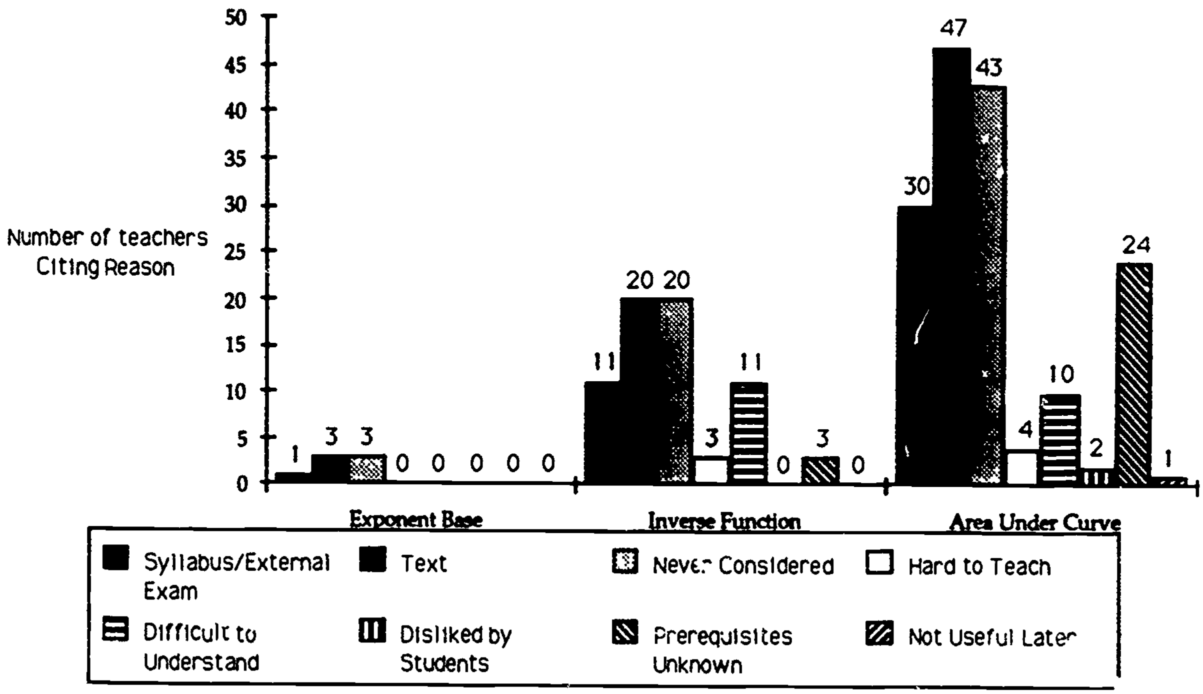
most frequently cited are: **text, never considered and syllabus/external examination.** As with the least used complex representation **(dilation, rotation)**, the least used logarithm representation **(area under curve)** had **prerequisites unknown** frequently cited.

Therefore, analogous to subtopic coverage, the data suggest that for a teacher to use a representation, not only must the representation be included in the syllabus or text, but the teacher must be familiar with the representation, the representation must be easy for students to understand, the students must know the prerequisites, and the teacher must know how the representation will be useful later. How much students enjoy a representation or how difficult it is to teach are much less important to teachers in selecting content representations.

Multiple Representations

When teacher coverage of logarithms and complex numbers is compared with the number of subtopics covered for each topic, numerous inconsistencies are found. For example, teachers did not mark a topic as covered even though they covered all four subtopics. Consequently, for a topic to have been covered either the teacher marked it as

Figure 9
Reasons Why Interpretation Was NOT Used



covered or the teacher covered at least three of the four subtopics. In our examination of multiple representations and their characteristics, we include only those teachers who covered the topics of complex numbers and logarithms.

Complex Numbers

The two multiple representations indices for complex numbers we examine are: (a) **Complex Frequent** and (b) **Complex Used**. For **Complex Frequent** we shall examine those teachers who frequently used: (a) at most one representation and (b) more than one representation. For **Complex Used** we shall examine teachers who used (either frequently or infrequently): (a) at most one representation, (2) exactly two representations, and (c) more than two representations Table 1 lists the numbers of each.

Table 1

Numbers of Teachers Using Multiple Representations

	Complex Frequent	Logarithm Frequent
Used at most one representation frequently	77	84
Used more than one representation frequently	44	32
	Complex Used	**Logarithm Used**
Used at most one representation	33	34
Used two representations	29	56
Used more than two representations	59	26
Total	121	116

Logarithms

For the complex number representations, teachers were specifically asked if they: (a) use this interpretation frequently; (b) have used this interpretation, but infrequently; or (c) do not use this interpretation. For logarithm interpretations, however, teachers were asked the number of *periods* they studied each interpretation. Therefore the construction of multiple representation indices for logarithms requires additional steps. First, class periods for each interpretation were converted into minutes which were then categorized as: not used, time = 0; used infrequently, 0 minutes < time < 75 minutes or about one period; and used frequently, time > 75 minutes or more than about one period. The results of this classification scheme are given in Table 2.

Table 2

Classification of Logarithm Representations

		Exponent Base	Inverse Function	Area Under Curve
Not Used (Time = 0)	5	35	84	
Used Infrequently (Time < 75 Minutes)	36	35	8	
Used Frequently (Time > 75 Minutes)		61	33	11
Total		104	103	103

While other classification schemes clearly exist, this scheme has three main advantages: (a) it is reasonable (There was a natural break in the data between 58 and 80 minutes for each interpretation.), (b) it allows us to compare and contrast multiple representation use for complex numbers and logarithms, and (c) it helps to eliminate complicating factors such as time spent on one particular representation.

Therefore, to parallel complex multiple representations, the two multiple representation indices for logarithms we shall examine are: (a) **Logarithm Frequent** and (b) **Logarithm Used**. For **Logarithm Frequent** we shall examine those teachers who frequently used: (a) at most one representation and (b) more than one representation. For **Logarithm Used** we shall examine teachers who used (either frequently or infrequently): (a) at most one representation, (2) exactly two representations, and (c) more than two representations. Table 1 lists the numbers of each.

Major Results

In this paper we examine only the major results of this analysis, that is, results that: (a) were supported by statistically significant relationships between at least two multiple representation indices and a variable <u>and</u> (b) had additional support from at least one other statistically significant relationship between at least one index and another closely related variable. Other results are given in a technical appendix (Glidden, in press).

Use and Development of Supplemental Materials

As Table 3 shows, there is a strong relationship ($p < 0.05$) between multiple representation use and the use of previously self-developed supplemental materials as sources of information on what to teach. As Table 4 shows, teachers who used multiple representations were also more likely to develop supplemental materials. This is especially surprising given how few teachers developed materials at all. Therefore, it appears, teachers who use multiple representations are more likely to develop and use supplemental materials.

Table 3

Relation Between Complex Indices and Sources of Information about Goals and What Topics to Teach is Materials Previously Prepared by Yourself

	Never Used	Occasionally Used	Frequently Used
Complex Frequent[a]			
Frequent <= 120	40	12	
Frequent > 1 5	22	14	
Complex Used[b]			
Used <= 1	10	19	1
Used > 2	1	20	5
Used > 2	14	23	20

Note. [a] C^2 (2, N = 113) = [illegible] $p < 0.05$
[b] C^2 (4, N = 113) = 19.492, $p < 0.001$

Content Presentation

As Table 5 illustrates, multiple representation teachers were more likely to use a minimum competency statement as a source of information on how to present a topic. Multiple representation teachers also were more likely to use the syllabus (or curriculum guide) and textbook (See Table 6.) as sources of ideas for problems that go beyond drill and practice. When Tables 3, 5, and 6 are viewed together, it is apparent that multiple representation teachers are more likely to use various resources (self-developed materials, minimum competency statement, text, or syllabus) for ideas than are single representation teachers.

Table 4

Relation Between Complex Used and Development of Supplementary Materials

	Developed Supplementary Materials No	Yes
Complex Used		
Used <= 1	33	0
Used = 2	19	10
Used > 2	41	18

Note. C^2 (2, N = 121) = 13.834, p < 0.001

Table 5

Relation Between Indices and Source of Information on How to Present a Topic is Statement of Minimal Competence

	Never Used	Occasionally Used	Frequently Used
Complex Frequent[a]			
Frequent <= 143	20	10	
Frequent > 1 13	16	12	
Complex Used[b]			
Used <= 1	16	9	6
Used > 2	19	5	2
Used > 2	21	22	14
Logarithm Frequent[c]			
Frequent <= 144	22	10	
Frequent > 1 8	13	10	

Note. [a]C^2 (2, N = 114) = 8.375, p < 0.05
[b]C^2 (4, N = 114) = 9.667, p < 0.05
[c]C^2 (2, N = 107) = 10.098, p < 0.01

Table 6

Relation Between Complex Indices and as a Sources of Information on Selecting Problems (e.g., applications) that go Beyond Drill and Practice

	Never Used	Occasionally Used	Frequently Used
	Syllabus or Curriculum Guide (Other than Minimum Comp ency Statement)		
Complex Frequent[a]			
Frequent <= 129	34	9	
Frequent > 1 7	23	12	
Complex Used[b]			
Used <= 1	15	11	4
Used > 2	10	14	3
Used > 2	11	32	14
	Textbook		
Logarithm Frequent[c]			
Frequent <= 136	29	10	
Frequent > 1 8	22	2	

Note. [a]C^2 (2, N = 114) = 8.704, $p < 0.05$

[b]C^2 (4, N = 114) = 10.087, $p < 0.05$

[c]C^2 (2, N = 107) = 8.148, $p < 0.05$

A strong relationship was found between the complex multiple representation indices and number of minutes spent on complex numbers. (See Table 7.) This is especially noteworthy when we recall that the majority of teachers used one logarithm representation and the time used for that interpretation varied from zero to 464 minutes. Therefore we would not expect our logarithm multiple representation indices to capture this relationship.

As already noted, there are major differences between teachers in time allotted. But as Table 8 illustrates, multiple representation teachers are not only more likely to cover more theorems, but they also are more likely to give more formal proofs. Therefore, there is strong evidence that multiple representation teachers spend more time on a topic and cover the topic more extensively than do nonmultiple representation teachers.

Table 7

Relation Between Complex Indices and Number of Minutes Spent on Complex Numbers

	Minutes		
	Mins < 180	180 <= Mins < 360	360 <= Mins
Complex Frequent[a]			
Frequent <= 1	15	28	24
Frequent > 1	4	10	28
Complex Used[b]			
Used <= 1	11	8	5
Used = 2	3	16	10
Used > 2	5	14	37

Note. [a]C^2 (2, $\underline{N}$ = 109) = 9.994, $\underline{p} < 0.01$
[b]C^2 (4, $\underline{N}$ = 109) = 27.930, $\underline{p} < 0.001$

Table 8

Relationship between Complex Indices and Initial Teacher Presentation

	Gave Formal Proof	Stated, Informal Derivation	Stated, No Derivation	Not Covered or Not Discussed
Formula $F(a + bi, c + di) = F(ac + bd, c^2 + d^2) + F(bc - ad, c^2 + d^2)\,i$				
Complex Frequent[a]				
Frequent <= 1	17	21	10	24
Frequent > 1	21	16	3	4
Complex Used[b]				
Used <= 1	2	6	4	16
Used = 2	8	8	5	8
Used > 2	28	23	4	4
Formula $\{r(\cos Q + i \sin Q)\}^n = r^n (\cos nQ + i \sin nQ)$				
Complex Used[c]				
Used <= 1	2	6	5	16
Used = 2	15	5	4	5
Used > 2	29	11	7	12

Note. [a]C^2 (3, $\underline{N}$ = 116) = 13.160, $p < 0.005$

[b]C^2 (6, $\underline{N}$ = 116) = 34.151, $p < 0.001$

[c]C^2 (6, $\underline{N}$ = 117) = 20.852, $p < 0.005$

Finally, as Tables 9 and 10 illustrate, there is evidence that teachers who use multiple representations for complex numbers are also likely to use multiple representations for logarithms.

Table 9

Relation between Complex Indices and Logarithm Frequent

		Logarithm Frequent	
		Frequent <= 1	Frequent > 1
Complex Frequent[a]			
Frequent <= 1	47	13	
Frequent > 1	20	16	
Complex Used[b]			
Used <= 1		24	3
Used = 2		18	8
Used > 2		25	18

Note. [a]C^2 (1, N = 96) = 5.537, $p < 0.05$

[b]C^2 (2, N = 96) = 7.444, $p < 0.05$

Table 10

Relation Between Complex Used and Logarithm Used

	Logarithm Used		
	Used <= 1	Used = 2	Used > 2
Complex Used			
Used <= 1	9	12	6
Used = 2	12	11	3
Used > 2	5	28	10

Note. C^2 (4, N = 96) = 11.029, $p < 0.05$

Teacher Experience and Education.

As Table 11 indicates, there is a strong direct relationship between experience in teaching mathematics and use of multiple representation. Additionally, Table 12 illustrates a statistically significant relationship between **Complex Frequent** and the number of semesters of mathematics methods. A similar, but not statistically significant, relationship is present between age and **Complex Frequent** and **Complex Used.**

Table 11

Relation between Indices and Number of Years Experience in Teaching Mathematics

		Yrs < 5	5 <= Yrs < 13	13 <= Yrs
Complex Frequent[a]				
Frequent <= 1	22	28	21	
Frequent > 1	11	9	23	
Logarithm Used[b]				
Used <= 1		13	9	9
Used = 2		14	22	17
Used > 2		3	4	17

Note. $^{a}C^2$ (2, N = 114) = 7.063, $p < 0.05$

$^{b}C^2$ (4, N = 108) = 15.090, $p < 0.05$

Table 12

Relation between Complex Frequent and Number of Semesters of Mathematics Methods and Pedagogy

		Semesters < 3	3 <= Semesters
Complex Frequent			
Frequent <= 1	37	34	
Frequent > 1	12	31	

Note. C^2 (1, N = 114) = 6.403, $p < 0.05$

Table 13

Relation between Logarithm Frequent and Head of Department

Logarithm Frequent	Head of Department Yes	No
Freq <= 1	37	39
Freq > 1	8	22

Note. C^2 (1, N = 106) = 4.268, $p < 0.05$

However, because of the results shown in Table 13 we cannot assert that there is a strong relationship between teacher experience/education and multiple representation use. In Table 13, **Logarithm Frequent** is inversely related to Head of Department. This could be attributed to the construction of the **Logarithm Frequent** index or possibly even to chance. But, the data also show relationships ($p < 0.05$) between head of department and age, experience teaching, experience teaching mathematics, and general education courses. Therefore, until further analysis is performed, we shall say that there is evidence of a relationship between teacher experience/education and multiple representation use.

Summary

External reasons (text and **syllabus/external examination**) and teacher familiarity (**well known** vs. **never considered**) frequently were cited as reasons for and against teaching particular subtopics of complex numbers and logarithms. Additionally, content reasons (**related to prior** and **useful later**) frequently are cited as reasons why a subtopic should be taught. Closely paralleling reasons for subtopic coverage, external reasons and teacher familiarity frequently were cited as reasons for and against using a particular concept representation and content reasons frequently were cited as reasons why a representation should be used. However, only for concept representation, **easy to understand** also was frequently cited as a reason for using a particular representation. For both subtopic coverage and concept representation, **easy to teach** and **enjoyed by students** were not often cited as reasons, either pro or con.

There were significant relationships between the use of multiple representations and teacher development and use of supplemental materials. There also was a relationship between multiple representation use and sources of information used to decide what to teach, how to teach, and what applications to present. Together these relationships suggest that teachers who use multiple representations use more sources of information (self-developed materials, minimum competency statement, text, or syllabus) than do nonmultiple representation teachers.

Teachers who use multiple representations also allot more time for a topic and they are more likely to cover important formulas and theorems more deeply than nonmultiple representation teachers. There was some evidence of a relationship between teacher experience/education and multiple representation use, but further research is necessary before inferences can be made.

Possible Implications for Mathematics Education

The results discussed above, if supported by further research, suggest several obvious implications for mathematics education regarding: development and use of supplemental materials, the relationship between sources of information and multiple representation use, and time allotted for coverage and depth of coverage. However, there is one less obvious implication that directly affects teacher education.

Recall Sherlock Holmes's "curious incident" of the dog not barking in "Silver Blaze" (Doyle, 1893). The fact that the dog did not bark was an important clue because it suggested that the dog knew the culprit. With respect to classroom process data, this analysis found no major relationships between multiple representation use and school data. Our data did not bark. Therefore, it appears that representation use is a local phenomenon, a function of teacher *perception*. That is, how familiar the teacher is with a representation, how easy it is for students to understand, how it relates to prior mathematics, and how useful it is for future mathematics. This perception may be influenced by the teacher's educational preparation and experience. This suggests that curriculum designers, supervisors, and mathematics educators should take special care to provide teachers with sufficient explanation of and justification for important concepts and their representations. Teachers make informed judgements regarding representation use (and subtopic coverage) and mathematics educators should be aware of this.

References

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*, New York: Holt, Rhinehart and Winston.

Doyle. A. C. (1893). *Memoirs of Sherlock Holmes*. New York: Harper.

Glidden, P. L. (In press), *Content representation in college algebra: Technical appendix*. Champaign, IL: International Association for the Evaluation of Educational Achievement.

Lebowitz, M. (1983). Generalization from natural language text.*Cognitive Science, 7*, 1-40.

McKnight, C. C., & Cooney, T. J. (in press). Content representation in mathematics instruction: Characteristics, determinants, and effectiveness." in L. Burstein (ed.), *Student growth and classroom processes in the lower secondary school: Vol. III. Second international mathematics study international reports*. London: Pergamon Press.

Novak, J. D. (1977). *A theory of education*. Ithaca, NY: Cornell University Press.

Piaget, J. (1975). *The development of thought*, New York: The Viking Press.

Second International Mathematics Study (SIMS) (1985). *Technical report V instrument book: Classroom process questionnaires*. Champaign, IL: International Association for the Evaluation of Educational Achievement.

Winston, P. H. (1972). Learning structural descriptions from examples. In P. H. Winston (Ed.) *The Psychology of Computer Vision*. New York: McGraw-Hill.

Appendix I

Short and Long Subtopic Titles

Logarithm Subtopics

Short Title	Long Title
Laws of Logs	Laws of Logarithms
Graphing Logs	Graphing Logarithmic Functions
Natural Logs	Natural Logarithms
Log Applications	Applications of Logarithms

Complex Number Subtopics

Short Title	Long Title
Complex Roots	Complex Roots of Quadratic Equations
Complex on Rect	Graphing Complex Numbers on Rectangular Coordinates
Complex with Polar	Polar Coordinate Representation for Complex Numbers
DeMoivre's Theorem	DeMoivre's Theorem and Roots of Unity

Similar issues in constructing explanatory indices from descriptive data are discussed in McKnight and Cooney, 1988, p. 4.

There also is a statistically significant relationship between teacher expectation of student mastery of $\log_b x = y$ and $\log_a x = z$ iff $b^y = a^z$ and Logarithm Frequent. However, because slightly more than 10% of the teachers did not teach the formula, the finding was not included. The chi-square statistic was significant at the 5% level for the relationships between teacher expectation of student mastery of the two formulas in Table 8 and Complex Used, but since several cells had expected counts less than 5, results could not be inferred from the data.
In fact, *only two* significant relationships were found between multiple representation use and all the school variables.
This is not to say that there may not be system differences in representation use. McKnight and Cooney (1988) found no clear, overall patterns of multiple representation use between systems, and there may be, and probably are, differences in preferred representations between systems. Further research is required to determine if a comparable implication can be inferred about other systems
There also was some evidence of a relationship between multiple representation use and the teacher's perception of class ability. These results are discussed the *Technical Appendix* .

CREATING GENDER DIFFERENCES:

A COMPARISON OF MALE AND FEMALE MATHEMATICS PERFORMANCE IN NINETEEN EDUCATIONAL SYSTEMS

Deborah Perkins Jones
David P. Baker

The Catholic University of America

INTRODUCTION

The search for differences between males and females across performance domains is a research activity undertaken by most disciplines in the behavioral and social sciences. To a great extent, differences (or similarities) between males and females are studied almost as by-products of phenomena in the pursuit of other theoretical interests -- such as cognition, physiology or social inequality. To a far lesser extent, differences between the sexes are studied as part of a theory of gender itself with an integrated set of hypotheses.

We examine several theoretical accounts of gender differences in one narrow performance domain. We compare, to the limits of our data, three accounts of gender influences from three broad segments of the sex difference literature -- sociological, social psychological and biological. The performance domain that we focus on -- eighth grade mathematics achievement -- is both narrow in content and short in the duration of an individual's life. But it is a domain that has demonstrated consequences for a range of behavior and later life chances.

We search for gender differences in mathematics achievement among 77,000 students within 19 educational systems around the world. We test the degree to which patterns of gender differences (or similarities) confirm central assumptions underlying each of the three theoretical perspectives.

THEORETICAL PERSPECTIVES ON GENDER DIFFERENCES

Since a measure of a subject's sex is easy to incorporate into most studies, a vast set of empirical findings about male and female differences has been produced. The same holds true for theoretical consideration of gender. Every conceivable theoretical perspective on human behavior contains an account of the origins of gender differences across an array of domains. The resulting literature is immense and unwieldy, without even the crudest of a central paradigm for conceptual guidance. This makes it difficult to place new evidence about males and females within a meaningful context. Faced with this kind of literature the task becomes one of theory reduction.

In generating hypotheses we have limited our consideration to three clusters of theoretical accounts of gender phenomena. These accounts represent central pools from which a large number of other theories flow. Also, each of these perspectives has some history of results in examining mathematical ability between the sexes. In our models of each perspective we do not claim to exhaust all of the numerous twists and turns of each theory, but rather we bring out data to bear on the central assumptions, the necessary conditions, of each of the general perspectives.

The Sociology of Gender Differences

Most sociological accounts of gender rest on the assumption that gender roles are born out of the institutions within a society. At the center of this idea is the notion that institutions define gender roles and that these definitions become forged into a diffuse "gender belief system" which shapes the day-to-day behaviors and attitudes of men and women, and girls and boys (Hess & Ferree, 1987).

By a sociological account then, the genesis of gender roles are the institutional rules of being a male or being a female. Other processes, more social psychological or even physical in nature, may transmit these rules to individuals, but at the heart of this perspective is the imagery of institutions forming rules about gender which in turn form the status of female or male within a society. A wide variety of institutions have gender-specific rules, such as rules of courtship and marriage, family organization, access to political power, and access and control over economic resources.

Related to this notion is that as institutional rules vary across societies, gender status varies across societies. Gender is considered to be actively and socially constructed; <u>it is not an immutable quality</u>. A central assumption of this sociological image is that differences in the relative status of the two sexes will correspond with the relative differences in performances. In societies in which there is a large difference between the status of men and women, there will also be large performance differences between men and women. In societies in which the relative status between the genders is small, performance differences between the genders should also be small. The reasoning behind this assumption being that gender will play less of a role in determining the conditions of performance of individuals in societies where gender is used less as a stratifying quality.

This is a main hypothesis of sociological explanations of gender differences in mathematical performance, but it has rarely been tested. A test of this hypothesis requires the kind of data we have -- namely, for a sample of societies a measure of sex differences in mathematical performance and measures of status differences between males and females. We have used a large cross-national data set on mathematical abilities of 8th graders as our indicator of gender difference in performance across societies. We have added to that a variety of measures of the relative status of men and women across a range of institutions. We have not attempted to form one global measure of gender status, but rather have selected indicators from several institutional dimensions. This allows us to assess the relative ability of various institutions to shape gender statuses which might create gender differences in performance.

Although we include indicators of general social status, we focus on economic indicators of gender status since differences between men's and women's access to financial resources and occupations seems to be a key correlate of a general gender status (Blau & Ferber, 1986; Chafetz & Dworkin, 1986). Also technical training and preparation for occupational positions are linked through attitudes towards formal schooling. Given the perception of mathematics training as an occupational skill, the basic sociological argument suggests that within a society with weak gender barriers to economic participation, there should be less gender differences in performance.

There is a related argument from the sociological perspective that we can examine. A number of global phenomena have resulted in limiting the degree to which social systems are structured (and stratified) by traditional attributes such as clan, family, ethnicity, and caste. The same case can be made for gender stratification as well.

The full host of influences on this process are too numerous to describe here, but the core of the argument usually centers on the phenomena of nation-state building and the process of creating citizens through formal schooling (Meyer & Hannan, 1979; Ramirez & Boli-Bennett, 1988). The argument goes that modern nation-states work to decrease traditional ties and increase citizen allegiance and participation. State sponsored institutions shoulder much of this task and chief among these institutions is formal schooling. It is suggested that schools through their state-derived charter and structure mitigate against traditional forms of stratification. As a student body and a future citizenry, children are less likely to be stratified in school by such qualities as their sex. The tone of this argument is essentially historical. As the Western model of states and schools spread, so did a decrease in the legitimate use of traditional mechanisms of stratification. The official implications of this trend can be seen in such governmental actions as the U.S. Title 9 prohibiting gender discrimination in school activities.

If this argument is true, we should find that gender differences in school performance decrease over time.[1] We can compare earlier national results of boys' and girls' performance on mathematics tests with the data we use to assess whether, as the conditions of gender stratification in a society decrease, so do gender differences in academic performance.

The Social Psychology of Gender Differences

There are a variety of social psychological accounts of gender differences. At the heart of most of these is the notion that face-to-face interactions in various social organizations influence the sexes in different ways thus yielding different performances. This basic scenario is very salient in the literature on gender and schools, in which a number of school factors are suspected of producing different experiences for males and females. These factors range from the imagery of a "hidden curriculum," which is thought to contain gender stratifying qualities, to more overt discrimination of access to educational opportunities (Becker, 1981; Brophy & Good, 1974; Fennema, et al., 1980; Leinhardt, Seewald & Engel., 1979; Morse & Handley, 1985).

The basic argument in all of these perspectives is that males are given advantages over females for the mastery of mathematics in school. And that these advantages are social psychological in nature, or namely effects of face-to-face interactions (Aiken, 1976; Burton, 1986; Walden & Walkerdine, 1986).

There are two such face-to-face schooling processes -- within classroom interaction and family effects -- which are often cited as causing gender stratification of performance. Research on the former considers how teachers might teach differently to

---

1 There are several other scenarios that generate essentially the same historical hypothesis about gender effects as does the nation-state and citizenship formation perspective. These include broader modernization arguments, arguments about the effects of conflict over traditional stratification and the expansion of industrial economies that break down traditional structures that incorporate gender heavily into the scheme of stratification. Our analysis tests only the basic hypothesis about the decline in gender effects on performance, not hypotheses about each of the numerous causal mechanisms that could play a part in this process.

male students and female students. And research on the latter considers how parents might influence their daughters and sons differently to achieve in school.

There is some cross-national evidence to suggest that males and females are universally treated differently in the schooling process (Finn, 1980). But also there is evidence to suggest the opposite, that formal education has become a force of gender egalitarianism, a place where females and males are treated similarly.

Whether or not males and females are taught differently is a large question that can not be completely answered with just one study. The data we use certainly does not contain measures of all possible gender discrimination that could occur during teaching in the classroom. It does, however, contain a measure of perhaps the most central of schooling processes determining performance, namely access to curriculum, or a student's "opportunity to learn" (OTL). We can determine if there are systematic gender differences in the opportunity to learn mathematics in these 19 educational systems. Do males gain an advantage in mathematics by being in classrooms where more and more advanced mathematics is taught? Or conversely, are females at a disadvantage because they are funnelled into classes where less and less advanced mathematics is taught?

Family influences as a possible explanation of gender differences in mathematics have been considered from a variety of perspectives, such as early socialization, forming performance expectations and standards, modeling of behavior conducive to solving mathematical problems and social reinforcements (e.g., Baker & Entwistle, 1987; Fox, Tobin & Brody, 1979). As is the case in all social psychological accounts of gender differences, the family is suspected of treating sons and daughters differently in terms of instilling the necessary skills to do mathematics (Eccles & Jacobs, 1986).

Since we do not have either direct family observation of parent-child interaction or parents' perceptions of their support, we do not focus on family effects in considering these social psychological arguments. We can however, include some investigation of the student's perceptions of their parents' encouragement to do well in mathematics and the student's attitudes about gender and mathematical training. We can assess the size of gender differences in these perceptions and attitudes and the relationship between these differences and performance differences across educational systems.

We test two central hypotheses of a social psychological perspective. First we can examine one "hidden curriculum" hypothesis, namely that boys receive more access to mathematical instruction than do girls and that this is a uniform pattern across educational systems. Second we can examine a more general socialization hypothesis, namely that parents encourage their sons' mathematical achievement more than their daughters' mathematical achievement and that this is a uniform pattern across educational systems.

The Biology of Gender Differences

Biological explanations of gender differences in mathematics achievement are ancient and varied, with the earliest speculation about cognition and gender differences dating back to Aristotle (Sherman, 1978). Current biological explanations reflect current core paradigms of biological thinking about an array of human performance, with accounts based on hormonal (Broverman, Klaiber, Kobayashi & Vogel., 1968) genetic (Bock & Kolakowski, 1973; Stafford, 1961) and neural structural effects (Levy, 1976; Waber, 1979).

Most biological theories rely on the relationship between spatial and mathematical ability. These theories argue that some biological characteristic (hormones, genes or brain structure) produces different degrees of spatial perception powers and this causes performance differences in solving mathematical problems. For the most part, these theories and the research that they spawn are relatively inductive.[2] They first assume that there are clear and consistent gender differences in solving mathematics problems and that the problem is to identify which biological factors, that are known to be distributed by gender, might account for the observed pattern of performance. Seldom, if ever, are the operative factors actually measured and tested against performance. This is partially because of the difficulty in measuring these factors, but equally it is because of the confidence in the inductive process behind much of this perspective. Consider, for example, Benbow and Stanleys' (1980) highly publicized paper in *Science*. They claim that since American male junior-high students out perform female junior-high students on a difficult mathematical test, males must have superior mathematical ability, "which may in turn be related to greater male ability in spatial skills" (p. 1264). They make this claim with only the scantiest of evidence about the lack of other non-biological effects at work within their data and without any

2 See Star (1978) for a similar critique of research on gender differences and brain hemisphere asymmetry.

direct measure of spatial skills. Their original claim may be correct, but they have not attempted to consider the extent to which biological and non-biological factors may shape the gender differences they observed.

Our data do not contain measures of operative factors from any of the biological theories of gender and mathematical performance. But with this data set we can extensively examine the core assumption behind the inductive chain of reasoning in these theories -- namely, are there consistent and large gender differences in mathematical achievement across a sizable number of students from different educational systems in different societies?

The degree to which the answer to this question is no, suggests a difficult obstacle for a general biological perspective on gender differences. A lack of consistent differences is not in and of itself a complete rejection of biological effects, since there are any number of genotypic and phenotypic analogies to suggest that biological influences can be masked by environmental ones. But at the very least, a lack of consistent differences would question the inductive reasoning that seems to buttress so much of the biological research about these phenomena.

Additionally a mixed pattern of gender differences would indicate the size of non-biological influences in these distributions. Short of offering some theory of societal influences on biological factors, an inconsistent pattern of effects suggests a variety of social influences.

Gender Difference as the Dependent Variable

Although most of the sex difference literature discusses phenomena in terms of individual differences between males and females, they are really investigating qualities of distributions. Except for a few gross anatomical characteristics, there is no evidence to suggest that all females differ from all males on any dimension. What we actually study is the distribution of one sex compared to the distribution of the other. We can examine how close the means are, or how spread out the distributions are relative to one another, and so forth. Thus we can then make probability statements about gender effects, such as "one's sex is likely to influence what one does or thinks or believes." These probabilities, however, are used merely to approximate individual qualities from aggregate qualities. The real currency of gender effects are differences or similarities among distributions of male and female performances. Therefore, we use comparisons of the male and female distribution of mathematical achievement from each system as our dependent variable.

Data and Measures

Data

The data on mathematical achievement come from the Second International Mathematics Study (SIMS) sponsored by the International Association for the Evaluation of Educational Achievement (IEA). SIMS, collected in 1981, is a comprehensive assessment of school mathematics of over 77,000 students in the grade equivalent to American 8th grade. Originally 20 national units participated in the study, these included: French Belgium (BFR), Flemish Belgium (BFL), British Columbia (BRC), England (ENG), Finland (FNL), France (FRA), Hong Kong (HKG), Hungary (HUN), Israel (ISR), Japan (JAP), Luxembourg (LUX), The Netherlands (NTH), New Zealand (NSL), Nigeria (NGR), Ontario (ONT), Scotland (SCT), Swaziland (SWZ), Sweden (SWD), Thailand (THA), and the United States (USA).[3]

The units do not represent a random sample of all nations in the world, rather they chose to participate in the study and each had control over their sampling and administering of the study instruments. The sample, however, does represent a reasonable mixture of the world's nations, including developed and less-developed nations and nations from most geographical regions of the world.[4] The sample of nations also represents a diverse set of administrative educational practices (Stevenson & Baker, 1989).

In each unit, a stratified, random sample of classrooms was drawn to the specifications of the guidelines developed by an international committee (Garden, 1987). The goal was to generate a representative sample of 13-year old students and schools in each educational system. A common mathematics test, minimally adopted for each country, was administered to these sampled intact classrooms at the end of the school

---

3 We can analyze only 19 of these systems because the Flemish Belgium sample did not contain a way to match the student's gender to their test performance. For the two provincial systems in Canada and the United Kingdom, we attempted to use province-level indicators of female status where possible. Also the Japanese sample was of 7th grade students and the Nigerian sample was 9th grade students, since in both systems the national committees deemed that the test tapped the mathematics curriculum at these grade levels.

4 See Jones (1989) for a full description of the sample and the SIMS study.

year. The test was designed to tap a range of mathematics skills, including four specific skill areas and five substantive areas. The total test contained 190 items made up on one 40-item core test and five rotated forms containing the remainder of items. Both the core test and the forms have a similar mixture of items in terms of skill and substantive areas. Each student took the core test and one of the rotated forms. Since all students in the study received the identical 40 item core test, we use just these items in our analysis. Additionally each student was asked to complete a questionnaire inquiring about their gender, their attitudes towards mathematics and their perceptions of their parents involvement in their preparation for mathematics.

Teachers of the sample classrooms were also given a questionnaire about what and how they taught mathematics to the target classroom. For each item on the test, teachers were asked to report if they had taught the information needed to answer the test item. This is the so-called Opportunity to Learn measure. See Appendix A for educational system, student and classroom sample sizes.

For each of the national units in the sample we collected a number of measures of gender status, economic development and the size of the school system. These came from published sources including: United Nations Demographic Yearbooks, UNESCO Statistical Yearbooks and Population Reference Bureau publications (Kent, Huab & Osaki, 1985; Sivard, 1985).

Measures

The gender difference on the core mathematics test for each educational system were calculated as the male mean score minus the female mean score. The individual scores from which the means were constructed were calculated as follows. Each core test item was a multiple choice with five optional answers. A core test score was computed using an estimated number know equation (Gulliksen, 1950).

$$\text{Core Score} = S\,R - (S\,W/4)$$

where: $S\,R$ = number of items correct
$S\,W$ = number of items incorrect

This scoring corrects for any effects of guessing.

The OTL for each teacher was calculated for each item and was merged onto the student records which enabled us to examine student level gender differences in access to mathematics instruction. Some OTL analyses were done at the classroom level using just the teacher file which included the gender breakdown of the classroom.

We use four indicators of women's status in non-economic institutions in each country. These include:

Fertility Rate - measured as the average number of children a women would have during her lifetime at current birth rates.

Percent Female Use of Contraceptives - measured as women in marital or consensual union, aged 15-49, using modern methods as defined as the pill, IUD, sterilization, condom, diaphragm, foam and other barrier or chemical methods.

Percent of Females aged 15-19 Married.

Number of Females in the national Legislation - Women both elected and appointed to legislative bodies.

We use six indicators of female status in the labor force for each country:

Percent Female in the Labor Force - as a percent of total labor force.

Percent Female in the Industrial Sector of the Labor Force.

Percent Female in the Service Sector of the Labor Force.

Percent Female in the Agricultural Sector of the Labor Force.

Gender Occupational Segregation Index - the degree to which females and males are concentrated in separate occupations.

Ratio of Female to Male Earning - averaged over all jobs.

Results

Table 1 presents the sex differences for each educational system on the 40-item core test. In the third column are the differences themselves (the male mean minus the female mean). Standard biological accounts of gender differences and numerous empirical studies suggest that males will outperform females on mathematical tests (Aiken, 1976; Backman, 1972; Benbow & Stanley, 1980; Maccoby & Jacklin, 1974; Mullis, 1975). This is not the case in these data. Instead, the differences fall into three distinct categories. In the first category are seven systems in which males do better than females. In the second category are eight systems in which there is no significant difference between the sexes. And in the third category are four systems in which females do better than males. There is also no evidence to suggest that the absolute size of a gender difference favors either sex. The absolute mean differences among the systems in the first group is 1.50 and in the third group it is an almost identical 1.51. The very small country-level mean difference of .30 reflects this mixture of gender differences across the systems. Also, a country-level mean difference weighted by the sample sizes shows less than a one-half of one item advantage for males (X = .49). Finally, it appears that mathematics performance is stratified less by gender than by educational systems. Here, as in earlier comparative mathematics studies (Husén, 1967), between-system differences are substantially larger than within-gender differences in any one system. These analyses offer little support for theories of gender that assume a consistent and uniform pattern of performance differences between the sexes.

Some of the more recent biologically grounded investigations of gender differences, however, have suggested that uniform differences will be most prevalent among the most difficult of mathematical areas. This is the male advantage hypothesis on so-called "higher order thinking" (HOT) involving spatial relationships, encouraged by the results that Benbow and Stanley (1980; 1983) report.

Although the core test was designed to tap a range of mathematics skills, we can examine the most difficult items to assess the HOT hypothesis comparatively. Within each system we determined the ten core test items which were most frequently answered incorrectly and then calculated the sex difference on these items for each system.[5] These differences are presented in the fourth column of Table 1.

In 12 out of the 19 systems the average male performed better than the average female on the 10 most difficult items on the core test. Also in no system did females significantly out perform males, as was the case with the full core test. This pattern lends some credibility to the hypothesis that males have an advantage in performing difficult mathematical problems, although the country-level un-weighted mean difference (.31) and the weighted mean difference (.38) are both relatively small. There were, however, seven systems in which this pattern did not hold and among these are

5 Because performance and the teaching of a mathematics curriculum varied so much between systems, we calculated the ten most difficult items within each system instead of across all systems.

Table 1

IEA Second International Mathematics Study 8th Grade Core Test Score Means by Gender.

Country	Male mean	Female mean	Difference (MX - FX)	Difference on 10 most difficult items (MX - FX)
I. $X_M > X_F$:				
France	17.02[a]	14.18	2.84*	1.00*
Israel	18.79	17.74	1.05*	.36*
Luxembourg	13.34	11.74	1.60*	.58*
The Netherlands	22.00	20.23	1.77*	.66*
New Zealand	14.60	13.51	1.09*	.69*
Ontario, Canada	17.72	16.94	.78*	.39*
Swaziland	9.29	7.89	1.40*	.26
II. $X_M = X_F$:				
British Columbia	19.55	19.27	.28	.41*
England/Wales UK	15.38	14.92	.46	.30*
Hong Kong	16.59	16.09	.50	.30*
Japan	23.84	23.80	.04	.31*
Nigeria	9.50	9.05	.45	.04
Scotland UK	16.83	16.68	.15	.24
Sweden	10.70	11.18	-.48	-.07
USA	14.98	15.12	-.14	.27*
III. $X_M < X_F$:				
Belgium-French	19.44	20.54	-1.10*	.13
Finland	13.24	14.87	-1.63*	.05
Hungary	22.36	23.62	-1.26*	-.01
Thailand	12.09	14.16	-2.07*	-.14
Country Mean (N=19)	16.17	15.87	.30	.31
Standard Deviation	4.24	4.37	1.24	.29

a. Scores on 40-Item Core Test were calculated as R - (W/4), where R is number of items correct and W is number of items incorrect.

* F ratio has $p \leq .01$

the four systems in which females did better than males on the full test. So there is countering evidence to suggest that even among the most difficult items there is not a uniform pattern of gender difference.

Before continuing the analysis we stop here to consider one major source of possible bias in these estimates. The SIMS sampling focused on the grade in which most 13 year-oldchildren were enrolled (i.e., the 8th grade). But not all of these educational systems are like the United States, in which nearly all 13 year-old males and females are in school in the same grade and thus yielding a nationally representative sample comparable for both sexes. If schooling in a particular system is selective for 13 year-old students and if this selection is somehow related to the gender of the student, then this could cause a biased comparison between male andfemale students from one grade level. While it is difficult to obtain precise estimates for each country of the percentages of the 13 year-old children, by gender, who are enrolled in the same grade level, we can make some rough estimates from which to judge any bias.

Fortunately most systems in the SIMS sample appear to be like the U.S. In only a few systems is there a possible comparison bias created by the structure and selectivity of schooling. These few cases are interesting to consider. Take for instance, France, in which there are substantially fewer males than females in the 8th grade school populations (and hence in the SIMS sample, see column 3 of Appendix A). This is due to a number of factors, chief among these is that over one half of French students repeat a year of school, and more boys than girls do this. In France repeating a year is often used as a proactive device to add an additional year of preparation for entrance to more difficult and prestigious technical secondary school streams (such as the "C-curriculum"), and boys apparently use this strategy more than do girls. Thus large male advantage in mathematics knowledge in the French sample may be upwardly biased, as we are comparing a smaller, slightly older and perhaps better prepared male population against a more general female population.

The reverse may be true in Nigeria in which there is low primary school enrollment in general (51% of an age-cohort in 1975) and male students outnumber female students by 2 to 1. The fact that we find no difference between the male and female means in mathematic ability probably underestimates male performance since we are comparing a broader population of Nigerian males against what is most likely a more selective group of Nigerian females.

By this type of reasoning, we estimate that some organizational bias may be involved in only five cases. The idiosyncratic structure of each case is too lengthy to describe here. We estimate, however, that among countries with a male advantage in mathematical achievement, certainly France and, to a lesser extent, Luxembourg and The Netherlands are upwardly biased. Among countries with parity between the sexes in achievement, Nigeria, as described above, may underestimate male performance. And among countries with a female advantage, Thailand is probably upwardly biased, but only to a small degree.

We next examine males' and females' access to mathematical instruction in the 8th grade. Table 2 presents the gender means and differences for OTL for the core test items in 14 of the systems.[6] The third column presents the gender differences in OTL. A central assumption of most social psychological accounts of gender differences in school settings suggests that through various mechanisms males have more access to mathematical instruction than do females, and this difference in access causes gender differences in performance. This assumption, however, does not receive much support from the gender differences in 8th grade OTL in the SIMS data. In fact, in one-half of the educational systems girls receive more mathematics instruction than do boys. And there is no difference in the full sample between male and female OTL means. There is also no correlation (r = .09) among the systems between gender differences in core test performance and gender differences in OTL. For example, among systems in which males perform better than females there is a mixed pattern of gender differences in OTL. Furthermore, in analysis not presented here, there is no evidence to suggest that 8th grade boys have more access to different or more difficult substance areas (arithmetic, geometry, algebra, measurement and probability) than do girls (Jones, 1989).

Although there appears not to be a male advantage in terms of access to instruction there may be other, subtle, ways in which one gender is given an advantage over the other. The so-called "hidden curriculum" perspective suggests that stratification within schools occurs through a variety of face-to-face mechanisms, some

6 Five systems did not collect OTL, but fortunately these systems are evenly distributed across the categories of gender differences, with one (Israel) from the male advantage category, two (England/Wales UK and Hong Kong) from the no difference category and one (Belgium-French) from the female advantage category.

very subtle and others more manifest. Since the students were sampled by intact classrooms, we can examine one such "hidden curriculum" hypothesis. Namely, that teachers alter the amount of mathematics they teach as a function of the gender composition of the class. This hypothesis flows from a number of "hidden curriculum" arguments which suggest that teachers, willingly or otherwise, take part in the social stratification of the schooling process.

The last column in Table 2 reports the unstandardized regression coefficient from regressing OTL on the percent female in the classroom. A negative coefficient indicates that teachers within a particular system decrease the amount of mathematics instruction as the number of female students increase. This is the case in only one system (The Netherlands). In the majority of systems the number of girls in a classroom has no effect on the amount of mathematics taught, and in three systems (the USA included) teachers teach more mathematics when there are more female students.[7]

Classrooms in these 19 educational systems appear to be generally equalitarian in terms of males' and females' access to 8th grade mathematics instruction. There is little support for the notion that schools manifestly limit classroom opportunities in 8th grade on the basis of the gender of the student. Finally what variation there is between gender and OTL is not related to the mixed pattern of gender differences in performance that we report in Table 1.

We next turn to several sociological explanations for the pattern of gender difference reported in Tables 1 and 2. A central notion of sociological perspectives on gender is that the relative status of men and women will influence sex differences in actions and attitudes. To the degree that a society's institutions create status differences between men and women, gender will be a stratifying characteristic. If this explanation

7 Carry the "hidden curriculum" notion further, one could argue that because female students tend to be better behaved in class (Entwisle & Hayduk, 1979), teachers with more female students can teach more of anything, mathematics included. So that the generally positive coefficients here do indeed represent a type of "hidden effect" of gender. To test this we added to the equations in Table 2 the teacher's estimate of the time spent on keeping order in this class. The effects of percent female were not diminished by adding this variable; teachers do not alter the amount of mathematics taught because more girls in class means better behaved students. If these positive coefficients represent a gender effect here, its underlying cause is not clear to us. We have also not separated out single-sex classrooms from this analysis, which could produce different gender effects from mixed-sex classrooms (Riordan, 1989).

Table 2

Access to Mathematics Instruction (OTL) in 8th Grade by Gender for Skills needed for Core Test Items.

	Country	Male mean % Core OTL	Female mean % Core OTL	Difference OTL Core (M - F)	Unstandardized coefficient from OTL regressed on % female in class (Standardized error)	
I.	The Netherlands	39.27	35.68	4.04*	-.05*	(.02)
	France	85.05	85.42	-.37*	.03	(.02)
	Ontario, Canada	80.79	80.76	.03	-.04	(.11)
	Luxembourg	59.24	61.27	-2.03**	.03	(.05)
	Swaziland	65.60	67.13	-1.53	.26	(.20)
	New Zealand	67.95	68.99	-1.04	.04	(.04)
II.	USA	77.50	78.82	-1.32**	.26**	(.08)
	British Columbia	25.55	27.07	-1.52**	.27*	(.13)
	Japan	81.43	81.43	0.00	.02	(.09)
	Nigeria	73.19	76.43	-3.25**	.07	(.09)
	Sweden	52.55	53.15	-.60	.09	(.06)
III.	Finland	63.15	64.07	-.92**	.14*	(.06)
	Hungary	48.61	49.34	-.73	.04	(.22)
	Thailand	84.84	84.46	.38	-.01	(.04)
	Country Mean (N=14)	65.6543	65.2871	-.6329		
	Country Standard Deviation	17.9766	18.3003	1.6384		

* Calculated F ratio has p≤ .05.
** calculated F ratio has p≤ .01

of gender phenomena is correct, we should find that the relative status of men and women will effect even a specific performance domain such as mathematics ability. To test this we examine the relationship between the relative status of females and males in a society and the size of the gender difference in 8th grade mathematics in the 19 educational systems in the SIMS data.

We begin with four indicators of general social status of females in a society. The correlations between these and the size of a society's gender difference in mathematics performance are presented in the first column of Panel A in Table 3. Contrary to the broadest interpretation of a sociological perspective on gender, general status of females is not related to the size of gender differences in performance of mathematics. Control over reproduction and marriage clearly are unrelated to performance differences. Political incorporation does show a modest association in the predicted direction, but the coefficient is not statistically significant.

In the first column of Panel B in this table we examine six variables that reflect various aspects of the integration of women into the institution of work. These indicators of the occupational status of females are related to the size of the gender difference in core test performance among the society's 8th grade students. In systems in which higher percentages of women work in the formal workforce, girls are more likely to perform as well or better than boys in mathematics. There appears to be a sector effect as well, with female participation in lower status agricultural work being less related to gender performance than female participation in higher status industrial work. Although the correlations for an index of occupational segregation and the ratio of female wages to male wages are not significant, each is in the predicted direction. All of these associations remain stable even after controlling for general economic development of the country (GNP) (analysis not reported here, see Jones, 1989) and many are statistically significant regardless of a small sample.

Table 3

Correlations Between Country-level Indications of Female Status and Gender Differences in 8th Grade on the Core Test, OTL and Attitudes.

	Gender Difference (M-F) on:			
	40-Item Core (N=19)	OTL (N=15)	Parental Encouragement (N=18)	Agree that Boys Need more Math (N=19)
Panel A:				
Women's Social Position				
Fertility Rate	.08	-.42	-.31	-.12
% Female Use Contraceptives	.02	.44	.09	-.84*
% Female 15-19 Married	-.14	-.49	-.19	.89**
# Female in National Legislation	-.32 (N=16)	.06	.30	-.25
Panel B				
Women's Labor Force Participation				
% Female Labor Force	-.55*	-.27	-.61**	-.20
% Female Industrial	-.59*	.01	-.42*	.24
% Female Service	-.40*	-.12	.12	.25
% Female Agricultural	-.24	-.28	-.42*	-.21
Gender Occ. Segregation	.33 (N=8)	.06	.30	.68* (N=8)
Female:Male Earnings	-.24 (N=11)	.18	-.15	-.47

* P ≥ .05.
** P ≥ .01

In the second column of Table 3 are the same correlations for gender differences in OTL. As we have already shown, there is considerably less range in gender differences in OTL than there is in performance and so we would not expect adult gender status to vary greatly with access to mathematics instruction. Unlike in the case of performance, gender differences in OTL are not associated with indicators of females' participation in the labor force. The indicators of female social status are also not related to gender differences in OTL.[8]

On a five point Likert Scale, with five as agreeing the most, there is considerable variation in the size of the gender difference on parental encouragement in mathematics. The overall mean is 1.4 (SD 5.9), with about a third of the countries having a female advantage, a third with parity between the sexes and a third with a male advantage. These system-level differences in perceptions of parental encouragement are associated with the performance gender differences (r = .47, p .02), so that countries that yield gender differences in performance also yield gender differences in perceptions of support by parents. These differences in parental support are also related to gender differences in OTL (r = .56, p .02). Table 3 shows that as with gender differences in performance, differences in parental encouragement are not related to indicators of general female status, but are related to indicators of female participation in the labor force. Systems in which females have more access to the labor force are systems in which there is less of a male advantage in parental support, and girls may even be more encouraged to do well in academic mathematics.

Differences between boys' and girls' agreement with the statement that "boys need mathematical training more than girls" are heavily in favor of males agreeing more than females with a sample mean difference of 10.6 (SD 11.5). The gender differences on this attitude are not related, however, to either differences in test performance or OTL (r = -.19 and r = -.14, respectively). And generally these differences are not related to female status, except for three indicators. In systems in which higher proportions of females use contraceptives, the gender differences in this attitude are smaller, but in systems with more young women marrying the gender difference in

8 The modest, although non-significant, correlations between OTL and fertility rate, contraceptive use and youth marriage are all in an unpredicted direction. This is largely due to the fact that The Netherlands, a country in which females have a more equal social status, has the largest male advantage in OTL and Nigeria, a country with considerably less parity between the sexes, has the largest female advantage in OTL.

favor of males agreeing is larger. Systems which yield higher levels of occupational sex segregation, also yield larger differences in the way boys and girls view gender and mathematics training.

The above analysis is cross-sectional. A related, but longitudinal, test of a gender sociological perspective on gender would suggest that the world over time creates social systems less structured around traditional attributes. Schools, and performance by male and female students within them, should reflect this change and thus gender difference in mathematics performance should decrease over time. To test this hypothesis we compared gender differences among 8th grade mathematics performance almost two decades apart in the nine countries that participated in both the First International Mathematics Study (FIMS) done in 1964 and the SIMS in 1981.[9]

The data presented in Table 4 supports the notion that there has been a decrease in the size of male superiority in 8th grade mathematics over the two decades. The sample mean drops from an almost 4% male advantage in 1964 to almost complete parity between the sexes in 1981. The individual country means show how this has happened. In 1964, all but one of the countries had a distinct male advantage mean difference. By 1981 four of these countries dropped substantially toward parity between the sexes. This trend has been noted in other data from just the USA (Kolata, 1989). Two countries (Belgium and Finland) actually replace a male advantage with a female advantage, a trend that runs counterto a strict interpretation of the hypothesis. Lastly, two countries have different patterns of means. Israel, the only country in 1964 with a female advantage, has a male advantage by 1981. And France's modest male advantage 20 years ago has strengthened over time.[10]

A further test of this notion of a decrease in gender differences in performance over time is to see if this is related to a change in the relative status of adult males and females over time. In other words, a general sociological perspective argues that as females gain more status relative to men in a society, performance differences between males and females decrease. We focus on changes in occupational status in all labor

9 The FIMS study was very similar to the SIMS in sampling, measurement and design. The core test was 30 items longer in the FIMS so we calculated a mean percent difference for each country.

10 In Israel this may be due to a sizable influx of Sephardic immigrants, since the early 1960s, who are more traditional in their use of gender as a stratifying quality. The results for France may be due to similar circumstances around increased immigration from Arabic societies.

sectors.[11] Figure 1 plots the change in female participation in the workforce from 1960 to 1980 against the change in gender differences in 8th grade mathematics from 1964 to 1981 for the nine countries that participated in both international studies.

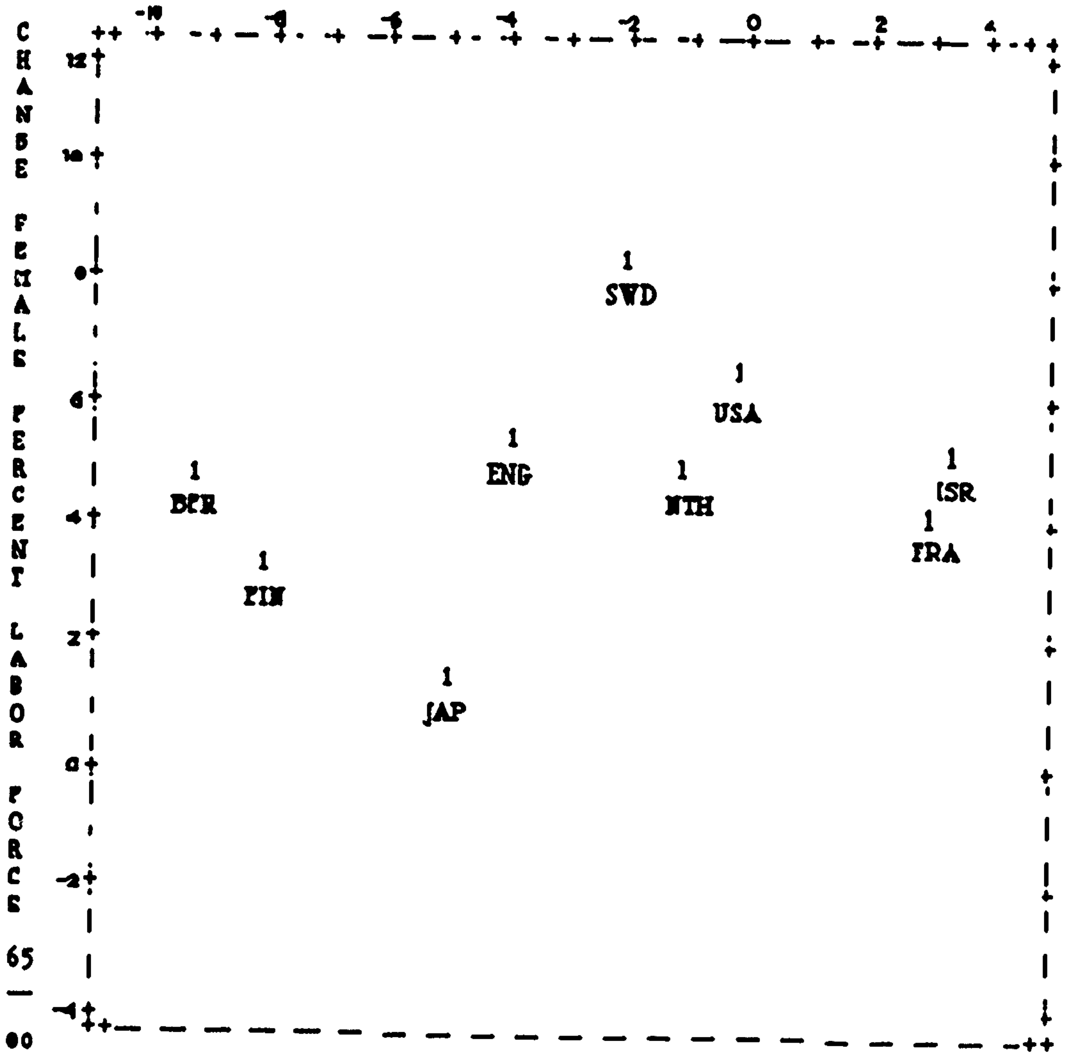
Most of the countries show the predicted relationship. Seven of the nine systems are in the upper most quadrant of the graphic with increases in female labor force participation associated with decreases in superior male mathematics performance over two decades. The relationship, however, appears not be strongly linear. The small sample of cases precludes standard tests of significance, but a non-parametric test of the ranking of the two variables yields a statistically significant relationship between the two variables. There are, however, several outlier in Figure 1 worth noting. First, Sweden shows less of a decrease in performance differences than its relatively large increase in a female labor force would predict. In part this may be due to the fact that the gender difference in 1964 was, like the U.S., already small. Secondly, both Israel and France go against the general trend by yielding an increasing male advantage in performance from 1964 to 1981.

Discussion

At the core of biological theories of gender performance differences on cognitive tasks are assumptions about universal and consistent gender differences in performance. The basic approach that is often used while testing biological hypotheses relies heavily on this assumption. Our results, however, provide little support for this central premise. Gender differences in 8th grade mathematics are not universal, nor are they uniform. There appears to be substantial variation by educational system as to the size and direction of gender differences. Many countries have no discernable differences and in countries with differences, males do not always have the superior performance. And, although there is some evidence to suggest that males do better than females on the hardest of mathematical problems, this tendency is not universal, as a sizable group of educational systems show no clear male advantage on these items.

11 We used various other sector combinations, such as non-agricultural, and found a similar pattern of results as those reported here for the more general indicator of female participation in all sectors of the economy.

Figure 1. Change in Female Labor Force Participation by Change in Gender Difference of 6th Grade Mathematics 1964-1981



9 cases plotted

Coefficient of Concordance (w) = .60

$\chi^2 = 5.44$ (df1), p< .01

$\bar{X}$ Female Labor Force Change 3.8 (3.2)
$\bar{X}$ Gender Diff. Core Test Change -3.03 (4.2)

TABLE 4

Comparison of Gender Differences in 8th Grade Mathematics Between the First International Mathematics Study (FIMS, 1964) and the Second International Mathematics Study (SIMS, 1981).

Country	Mean % difference Males - Females	
	1964[a]	1981[b]
Belgium[c]	6.43	-2.75
England/Wales	5.36	1.15
Finland	4.07	-4.10
France	4.29	7.10
Israel	-3.00	2.63
Japan	5.36	0.10
The Netherlands	6.64	4.43
Sweden	2.57	-0.01
United States	1.07	-0.35
Country Mean	3.94	0.91
Country Standard Deviation	2.39	3.45
T	2.15	
(df)	8	
One-tailed test p=	.032	

a. Adopted from Husén (1967, p. 240), percentage estimated number known of 70-item Core Test for population 1b.

b. Percentage estimated number known of 40-item Core Test.

c. Belgium sample in 1964 is from the entire country, and the 1981 sample used here is from the Belgium-French proportion of the country

Our analysis does not include information about the main operative factors in biological arguments on gender. We do not measure spatial ability, hormonal, genetic or neurological effects and our results are not incompatible with a biological interpretation that would suggest that underlying universal gender effects are masked or enhanced in various situations. Our results do suggest though a fundamental task for the biological approach. Namely, if there are biological effects, they must be measured directly and their size relative to non-biological must be assessed. It is not enough to assume that universal gender effects exist in mathematics performance at the 8th grade on the basis of results from students in one educational system. And it is not enough to merely search for biological factors that correlate with gender as an explanation for assumed performance differences. Until all of these pieces are pulled together into a unified approach we will know little about the existence of biological influences on the creation of gender differences in mathematical performance.

This variation in the pattern of gender difference suggest that there are sizable social influences in their creation. Our analysis has explored several explanations for these phenomena.

We have shown that schools are generally equalitarian in terms of boys' and girls' access to training in mathematics at the 8th grade level. Contrary to a central tenet of a social psychological approach, which suggests that the sexes are treated differently in school and that boys are often favored, boys do not receive more training in mathematics. And in some systems girls actually receive more training on the average than do boys. We do not have data on other central processes that make up the "hidden curriculum" perspective. For instance, we do not know if within classrooms, teachers teach differently to female students than they do to male students and so forth. Nor do we have measures on a host of other face-to-face processes which could be stratified by gender, such as the effects of guidance counselors for example (Fox, et al., 1979; Pietrofesa & Schlossberg, 1977; Shafer 1976). But to the degree that our extensive measure of OTL taps general access to mathematics, schools do not seem to favor boys by teaching them more mathematics than they teach girls.

Lastly, our analysis has yielded some evidence for a sociological perspective on gender differences, although the data suggest that the sociological process is not as general as it is often assumed to be. While societal level indicators of gender parity in the labor force are generally related to gender difference in 8th grade mathematics performance, other indicators of gender status are not related to performance differences. The effects of gender status on performance seems to be domain specific. If a society incorporates more women into the formal labor market, its students exhibit less gender differences in school mathematics performance. But if a society incorporates women into other domains, this may or may not effect gender differences in performance.

Gender status is not monolithic across all institutions within a society. Rules about gender vary across institutions within the same society; and the degree to which one institution is connected to another will shape how much or how little gender will play a part in the roles under joint control of these institutions.

Schooling and the labor market are strongly connected in most societies. To the degree that school is an institution of preparation for the work place, our findings verify a sociological creation of gender performance differences. In systems in which girls have more of an option to enter the labor force, their performance on mathematics is more similar to boys. The social process behind this phenomenon is hinted at through our analysis of the student's perception of parental encouragement to study mathematics. Systems with more gender parity in parental encouragement are those with more females in the labor market. Also gender parity in parental encouragement is related to gender parity in performance. In societies with labor force opportunities for both men and women, parents encourage both their sons and daughters to study mathematics and both boys and girls do this. These results suggest that social opportunities (or barriers to opportunity) resonant down to performances of actors within social systems.

Further we found that, as a general modernity hypothesis would predict, gender differences in mathematics performance have decreased over time in nine educational systems scattered around the world. This parallel earlier evidence to suggest that gender has become less of a barrier to access to mathematics and science instruction in 8th and 9th grade (Keeves, 1973).

The incorporation of women into a wider sphere of economic participation in many societies has been the result of a number of processes, chief among these being the expansion of schooling on a Western model and the state's breakdown of traditional modes of social stratification. Besides the economic benefits of this process (Benavot, 1989), the belief that its full adult population is a nation's chief economic resource has become a standard political notion. Witness the recent publicity about the "crisis" over the decline in American students who will enter scientific and technical training and the calls for the expansion of training in these areas to a full range of students (Johnston & Packer, 1987; Walker, 1988).

These kinds of processes seem to filter down to the interests and actions of individuals. The fact that most of the nine educational systems for which we had longitudinal data exhibited dramatic declines in male advantages in mathematics in less than 20 years indicates the potency of these social effects on individual behavior.

There are a number of reasons to be cautious about making too sweeping a conclusion from our results as to their bearing upon central assumptions of theories of gender effects on performance.

First, the SIMS data includes only one subject. Similar analyses should be done for other academic subjects. Particularly those subjects, such as reading, for which girls have been thought to have an inherent (or otherwise) advantage over boys (Maccoby & Jacklin, 1974). This would broaden a comparative treatment of gender effects. For example, a sociological argument would suggest that a decrease in gender stratification would also reduce gender differences in reading. Also, analysis of subjects that depend on mathematical ability, such as science, should be done to verify the findings we report for mathematics.

Second, we have concentrated only on 8th grade performance. A comparative analysis should be done on other school levels. This is particularly important for secondary schooling, for which a number of hypotheses exist about gender effects on curriculum tracking and choice of subjects that can influence performance factors.

Third, in examining the assumptions behind sociological accounts of gender we have focused on economic and general social statuses of women and men. Other institutions need to be considered. For example, within schooling itself, certain institutional arrangements can foster status parity or differences between the sexes which could be hypothesized to influence performance. These would include the relative opportunities for technical training for males and females later in school and so forth.

Fourth, although the SIMS sample of national educational systems is moderately large (about 10% of all nations in the world), the sample was not as representative of less developed countries and certain portions of the world (i.e., Latin and South America) as one would like. Gender effects may be different in these systems, although we do not know of any major arguments to suggest that these omissions would have greatly changed the overall pattern of results. Also in some of the analysis we were forced to use a reduced sample and thus had to give more weight to any outlying cases.

Additionally, the SIMS data does not contain all of the variables one would like to have to analyze most "hidden curriculum" arguments. We only had a classroom level estimation of access to instruction. Although this is important, there is other research to suggest that within classroom access can be stratified by gender (Hallinan & Sorensen, 1987). Also, the data set is not as sensitive to a number of within country variations from which one could pull collaborating evidence of the processes we have looked at here (Schildkamp-Kündiger, 1982; Theison, Achola & Boakari, 1983).

Aside from these caveats, these data and other IEA data sets are the best available to assess academic performance comparatively. The careful standardization of test items, the attempts to make each within country sample representative of schooling, and the overall size of the number of students, teachers, classrooms and schools involved lend credibility to any results derived from these data. Until there are better data, these represent the best estimate that we have on the relative effects of gender on similar mathematics tests around the world.

Conclusion

How well do the central assumptions of the three general perspectives on gender phenomena fare in light of our evidence on gender and mathematics performance across 19 educational systems? We find mixed evidence for all three perspectives, with some variation in the clarity of the evidence. Our findings are most damaging to the naivest of biological arguments and are most supportive of sociological perspectives if they are modified to consider specific institutional effects. Although there is clearly some kind of social psychological process at work here, a central assumption of a "hidden curriculum" perspective is not supported. We establish clear evidence of a world trend in schools to give females access to mathematical training at the 8th grade. Still, there are many unanswered questions within a general "hidden curriculum" approach to gender stratification within schools.

Perhaps most importantly, our results have demonstrated the advantages to considering gender effects comparatively. Without this kind of perspective it becomes very difficult to consider a full range of hypotheses. As we have shown, the lack of comparative data has led to the building of some theoretical perspectives on unwarranted assumptions about how females and males perform cognitive tasks.

Most interesting is the evidence, particularly the longitudinal results, suggesting that sociological processes may lessen gender effects on performance. Until now these processes have generally been left untested within the area of mathematical performance. Being in one society versus another has ramifications not only for the level of mathematics students master, but also for the level of gender stratification of that knowledge.

Male superiority in mathematics performance in schools has decreased over the last two decades, this trend seems to be related to the greater incorporation of women into the labor market. It may also be related t the even larger process of citizen formation and the incorporation of a modern notion of the individual (Ramerez & Boli-Bennet, 1988). This needs to be tested further.

This sociological evidence about the size and direction of gender differences in educational systems merits further comparative consideration. This evidence, we think, especially merits consideration by proponents of theories that assume that gender effects on performance are only created by face-to-face or biological processes.

REFERENCES

Aiken, L.R. (1976) Update on Attitudes and Other Affective Variables in Learning mathematics. Review of Educational Research, 46(2), 293-311.

Backman, M.E. (1972). Patterns of Mental Abilities: Ethnic, Socioeconomic, and Sex Differences. American Educational Research Journal, 9, 1-12.

Baker, D. & Entwistle, D. (1987), The Influence of Mothers on Academic Expectations of Young Children: A Longitudinal Study of How Gender Differences Arise. Social Forces, 65(3), 670-694.

Becker, J.R. (1981). Differential Treatment of Females and Males in Mathematics Classes. Journal for Research in Mathematics Education, 12(1), 40-53.

Benbow, C.P., & Stanley, J.C. (1980). Sex Differences in Mathematical Ability: Fact or Artifact? Science, 210, 1262-1264.

Benbow, C.P., & Stanley, J.C. (1983). . Sex Differences in Mathematical Reasoning Ability: More Facts. Science, 222, 1029-1031.

Benavot, A. (1989). Education, Gender, and Economic Development: A Cross-National Study. Sociology of Education, 62, 14-32.

Blau, F. & Ferber, M. (1986). *The Economics of Women, Men and Work.* Englewood Cliffs, NJ: Prentice-Hall.

Bock, R. & Kolakowski, D. (1973) Further Evidence of Sex-Linked Major-Gene Influence on Human Spatial Visualizing Ability. American Journal of Human Genetics, 25, 1-14.

Brophy, J.E. & Good, T.L. (1974). *Teacher-Student Relationships.* New York: Holt, Rinehart, & Winston.

Broverman, D.M., Klaiber, E.L., Kobayashi, Y., & Vogel, W.. (1968). Roles of Activation and Inhibition in Sex Differences in Cognitive Abilities. Psychological Review, 75(1), 23-50.

Chafetz, J. & Dworkin, A. (1986). *Female Revolt: Women in World and Historical Perspective.* Totowa, NJ: Rowman & Allanheld.

Eccles, J.S. & Jacobs, J.E. (1986). Social Forces Shape Math Attitudes and Performance. Signs: Journal of Women in Culture and Society, 11(2), 367-380.

Entwisle, D. & Hayduk, L. (198[illegible]). *Early Schooling.* Baltimore: Jones Hopkins.

Fennema, E. .(1980) Sex-related Differences in Mathematics Achievement: Where and Why? In L. Fox, *Women and the Mathematical Mystique.* Baltimore: Johns Hopkins.

Finn, J. (1980). Sex Difference in Educational Outcomes: A Cross-National Study. Sex Roles, 6(1), 9-28.

Fox, L.H., Tobin, D. & Brody, L. (1979). Sex-Role Socialization and Achievement in Mathematics. In M. Wittig & A. Petersen (Eds.), *Sex-Related Differences in Cognitive Functioning: Developmental Issues*. New York: Academic Press.

Garden, R.A. (1987). The Second IEA Mathematics Study. Comparative Education Review, 31(1), 47-68.

Gulliksen, H. (1950). *Theory of Mental Tests*. New York: John Wiley & Sons.

Hallinan, M. & Sorensen, A. (1987). Ability Grouping and Sex Differences in Mathematics Achievement. Sociology of Education, 60(2), 63-73.

Hess, B. & Ferree, M. (1987). Introduction. In B. Hess & M. Ferree (Eds.), *Analyzing Gender: A Handbook of Social Science Research*. Newbury Park, CA: Sage.

Husén, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*. Vols. I & II. Stockholm: Almqvist & Wiksell.

Johnston, W. & Packer, A. (1987). *Workforce 2000: Workers for the 21st Century*. Indianapolis: Hudson Institute.

Jones, D. (1989). Gender Difference in Mathematics Achievement: Across-National Analysis. Unpublished Doctoral Dissertation, The Catholic University of America.

Keeves, J. (1973). Differences Between the Sexes in Mathematics and Science Courses. International Review of Education, 19, 47-62.

Kent, M., Huab, C. & Osaki, K. (1985). *The World's Women: A Profile*. Washington, DC: Population Reference Bureau.

Kolata, G. (1989). Gender Gap in Standard Tests in Narrowing, Experts Find. The New York Times, July 1, 1.

Leinhardt, G., Seewald, A. M. & Engel, M. (1979). Learning What's Taught: Sex Differences in Instruction. Journal of Educational Psychology, 71(4), 432-439.

Levy, J. (1976). Cerebral Lateralization and Spatial Ability. Behavior Genetics, 6(2), 171-188.

Maccoby, E.E. & Jacklin, C.N. (1974). *The Psychology of Sex Differences*. Stanford, CA: Stanford University.

Meyer, J. & Hannan, M. (Eds). (1979). *National Development and the World System*. Chicago: University of Chicago Press.

Morse, L. & Handley, H. (1985). Listening to Adolescents: Gender Differences in Science, Classroom Interaction. In L. Wilkinson & C. Marrett (Eds.), *Gender Influences in Classroom Interaction*. Orlando: Academic, 37-56.

Mullis, I.V.S. (1975). *Educational Achievement and Sex Discrimination*. Denver: National Assessment of Educational Progress.

Pietrofesa, J. & Schlossberg, N. (1977). Counselor Bias and the Female Occupational Role. In J. Pottker & A. Fishel (Eds.), *Sex Bias in the Schools*. Teaneck, NJ: Fairleigh Dickenson University.

Ramirez, F. & Boli-Bennet, J. (1988). The Political Construction of Mass Schooling: European Origins of World Wide Institutionalization. Sociology of Education, 60(1), 2-17.

Schildkamp-Kündiger, E. (Ed.). (1982). *International Review on Gender and Mathematics*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics & Environmental Education.

Shafer, S. (1976). The Socialization of Girls in the Secondary Schools of England, and the Two Germanies. International Review of Education, 22, 5-23.

Sherman, J. (1978). *Sex Related Cognitive Differences: An Essay on Theory and Evidence*. Springfield, IL: Charles C. Thomas

Sivard, J. (1985). *Women...A World Survey*. Washington, DC: Population Reference Bureau.

Stafford, R.E. (1961). Sex Differences in Spatial Visualization as Evidence of Sex-Linked Inheritance. Perceptual and Motor Skills, 13, 428.

Star, L. (1979). Sex Differences and the Dichotomization of the Brain: Methods, Limits and Problems in Research on Consciousness. In R. Hubbard & M. Lowe (Eds.), *Genes and Gender*, Vol. II. New York: Gordian, 113-131.

Stevenson, D. & Baker, D. (1989). A Comparative Analysis of the Implementation of the Mathematics Curriculum. Unpublished paper, The Life Cycle Institute, Catholic University of America.

Theison, G. Achola, P., & Boakari, F. (1983). The Underachievement of Cross-National Studies of Achievement. Comparative Educational Review, 27, 46-68.

Waber, D.P. (1979). Cognitive Abilities and Sex-related Variations in the Maturation of Cerebral Cortical Functions. In M.A. Wittig & A.C. Petersen (Eds.), *Sex-related Differences in Cognitive Functioning*. New York: Academic Press.

Walden, R. & Walkerdine, V. (1986). *Girls and Mathematics: From Primary to Secondary Schooling*. London: Institute of Education, University of London.

Walker, W. (1988). Study Sees Big Leap in Science Jobs by 2000. Science, June, 13.

APPENDIX A

Sample Size for Students and Classrooms for the Nineteen SIMS Educational Systems.

Country	Student Sample	Classroom Sample	Male:Female Ratio Sample
BFR	2086	105	1.14
BRC	2567	92	0.98
ENG	2678	416	0.85
FNL	4484	206	1.10
FRA	8778	365	0.77
HKG	5548	130	1.03
HUN	1753	70	0.93
ISR	3819	153*	1.04
JAP	7785	211	1.06
LUX	2106	107	0.97
NGR	1465	46	2.68
NTH	5500	236	1.04
NZL	5978	196	1.02
ONT	6222	213	1.01
SCT	1356	354*	1.16
SWD	3585	186	1.10
SWZ	904	25	0.86
THA	4030	99	1.08
USA	6957	250	0.93
Total	77,602	2681	

*Number of teachers, not actual classroom count.

APPENDIX B

Means and Standard Deviation for Female Status Variables

Variable	Mean	Standard Deviation
Fertility Rate	2.4	1.4
% Female Use Contraceptives	68.2	18.3
% Female 15-19 Married	7.4	8.8
# Female in ational Legislation	11.4	9.7
% Female Labor Force (1960)	32.7	7.8
% Female Labor Force (1980)	36.6	6.4
% Female industrial	23.9	8.4
% Female Service	45.7	7.2
% Female Agricultural	31.1	10.8
Gender Occupation Segregation index	40.1	7.0
Ratio Female to Male Earnings	73.8	7.0

A MULTI-LEVEL MODEL OF SCHOOL EFFECTIVENESS IN A DEVELOPING COUNTRY

Marlaine E. Lockheed
The World Bank

Nicholas T. Longford
Educational Testing Service

INTRODUCTION

Although appropriate methods for analyzing hierarchically structured data have been available since the early 1970's (Dempster, Laird & Rubin, 1977; Lindley & Smith, 1972), application of these methods to educational policy decisions in developing countries has been hampered by two important shortcomings: (a) the absence of computationally efficient algorithms for multi-level analysis, and (b) the lack of adequate data (sufficient cases at each organizational level). Recently, new computational methods have been developed that address the first problem (Goldstein, 1986; Longford, 1987; Raudenbush & Bryk, 1986), and data sets sufficient for their application have been collected in a number of developing countries. This paper applies one of the techniques to longitudinal data recently collected by the International Association for the Assessment of Educational Achievement (IEA) in Thailand to answer three important questions for policy-makers: Which characteristics of schools and teachers are associated with student learning over time? To what extent? And, are the differences among schools uniform across different types of students, or are some schools more effective with certain types of students?

The comparative effectiveness of schools, particularly the relative efficiency with which alternative inputs and management practices enhance student achievement, has become the center of a lively debate in the literature (see, for example, Goldstein, 1984; Heyneman, 1986; Reynolds, 1985; Rutter, 1983; Willms, 1987). These issues have important implications for how governments and international development agencies should allocate their limited resources—whether they should concentrate on certain types of inputs (capital investment, lowering class size) or should finance others (instructional materials, teacher or headmaster training, student testing). In the United States and United Kingdom, the debate was sparked by studies that claimed to identify effective schools: those that enhanced student achievement more than other schools working with similar students and material inputs (see Raudenbush, 1987, for a recent review). In developing countries, research on school effectiveness has been limited; studies that have examined the effects of alternative inputs on student achievement have not taken into account the explicitly hierarchical nature of the explanatory models and data.

The "effective schools" issue has been fueled by controversy over methodology, interpretation and data (for example, Sirotnik & Burstein, 1985). The most important methodological issue is the use of inappropriate statistical models for analyzing multi-level data. The argument concerns how behavior at one level (e.g., classroom, school, district) influences behavior at a different level (e.g., students), and how to correctly estimate these multi-level effects.[1] Hierarchically structured data are common in social research, because social institutions are typically hierarchically organized, but commonly used statistical techniques for dealing with related data may lead to biased estimates.[2] In particular, it has been established that, when observations within clusters on any stratum are more homogeneous than those between clusters, using ordinary least squares (OLS) regressions with such data can lead to biased estimates of regression coefficients in unbalanced designs, and to substantially biased standard errors for these estimates even in balanced designs. Most policy research entails the use of unbalanced designs, and so a serious problem may arise when ordinary least squares regression estimates are used for quantifying the effects from alternative inputs.

Proper analysis of multi-level data entails two distinct changes in thinking about data. First, the demands of inherently hierarchical data, such as much education data, need to be confronted at the conceptualization stage, so that sufficient numbers of units at each level are sampled (e.g., adequate samples of schools and classrooms, in addition to sampling of students). Second, and more important, hierarchical analysis requires a major shift in how problems of organizational effects on individuals are viewed; instead of considering only effects of <u>levels</u>, effects on <u>relationships</u> are also modelled. For example, in education, certain school or classroom interventions may affect not only average student achievement, but also lessen hypothesized correspondence between family background and student achievement. Here are organization-level force serves to mediate individual-level effect.

Until recently, most discussions of multi-level analysis have remained theoretical, bounded by costs and computational requirements of existing analytic tools. However, the debate has been energized by the recent development of new analytic tools for analyzing multi-level data (Aitkin & Longford, 1986; Goldstein, 1986; Mason, Wong & Entwisle, 1984; and Raudenbush & Bryk, 1986). Although the development of the general EM algorithm (Dempster, Laird & Rubin, 1977) provided a theoretically satisfactory and computationally manageable approach to covariance component estimation in hierarchical linear models, it has seen limited application in education policy research due to three shortcomings: slow convergence of the algorithm, lack of suitable generally available software, and lack of understanding of these techniques in the education research community. The new tools, by comparison, offer computational algorithms for variance component analysis of hierarchically structured data that converge rapidly and require only a moderate amount of computation in each iteration. The research described here utilizes the software VARCL which implements the Fisher scoring algorithm of Longford (1987) to address important policy questions regarding effectiveness and efficiency of education in developing countries.

To date, application of the new tools in education policy research has been limited to relatively few studies of schools in developed countries; to the best of our knowledge, this is the first such application to data from developing countries. Other research on developing countries has demonstrated that school-level inputs have significant effects on student achievement (for example, Fuller, 1987; Heyneman & Loxley, 1983; Heyneman & Jamison, 1980; Lockheed & Hanushek, 1988; Psacharopoulos & Loxley, 1986). However, previously employed analyses have not addressed the problem of multi-level data and may have over- or underestimated the importance of classroom, school and district-level effects, which are those that governments and donors can best address.

DESIGN

Analytical Framework

This project makes an important methodological contribution by application of multi-level models to estimation of school and classroom effects on student achievement. The problem with ordinary least squares (OLS) estimates of school and classroom effects have been discussed at length by Aitkin & Longford (1986) and Dempster, Rubin & Tsutakawa (1984). In short, these problems arise from the nature of typical data in educational surveys.

Educational surveys involve hierarchically structured data—pupils within classrooms, within schools, within administrative units or regions. Every classroom (school, region) has its own idiosyncratic features that result from a complex of influences, including composition, teaching practices and management decisions.

As a consequence, observations on students (e.g., their outcomes) are not statistically independent, not even after taking account of the available explanatory variables. This presents a violation of the assumptions for ordinary regression (OLS). The main problem is not so much with the estimates themselves as with their standard errors, and adjustment techniques based on the "design effect" are not satisfactory for complex regression models.

Variance component models are an extension of ordinary regression models; the extension refers to more flexible modelling of the variation. Pupils are associated with (unexplained) variation, but this variation has a consistent within-classroom component, which itself has a within-school component, etc. Schools vary, classrooms within schools vary and pupils within classrooms vary.

Consider the regression models for data with two levels of hierarchy (pupils j within classrooms i):

$$(1)\quad y_{ij} = a + bx_{ij} + cz_{ij} + e_{ij}$$

where a, b, c are (unknown) regression parameters, x and z are explanatory variables, y the outcome measure and the random term e is assumed to be a random sample from $N(O, s^2)$. Variation among the classrooms can be accommodated in the "simple" variance component model

$$(2)\quad y_{ij} = a + bx_{ij} + cz_{ij} + a_i + e_{ij}$$

where the a's form a random sample (i.i.d.) from $N(O, t^2)$ and the a's and the e's are mutually independent. The covariance of two pupils within a classroom is t2 (correlation $t^2/[r2 + s^2]$). If we knew the a's we could use them to rank the classrooms. The model (2) has the form of analysis of variance (ANOVA), with distributional assumptions imposed on the a's. The advantages of this assumption are discussed by Dempster, Rubin and Tsutakawa (1981) and Aitkin and Longford (1986). In the former reference the term "borrowing strength" in estimation of the effects of small groups is used. In addition, some schools may be more "suitable" for pupils with certain backgrounds than others. This corresponds to variation in the within-school regressions of y on x and z, and this situation can be suitably modelled as

$$y_{ij} = a + bx_{ij} + cz_{ij} + a_i + b_ix_{ij} + t_iz_{ij} + e_{ij},$$

or

$$(3)\quad y_{ij} = a + bx_{ij} + cz_{ij} + a_i + b_ix_{ij} + e_{ij}.$$

The classroom-level random effects (a_i, b_i) are assumed to be a random sample from $N_2(0, S_2)$; here S_2 involves only 3 parameters: the variances of a and b and their covariance. Extensions to larger numbers of explanatory variables and to more complex hierarchies are described in the literature (e.g., Goldstein, 1987; Longford, 1987; Raudenbush & Bryk, 1986).

The maximum likelihood estimation procedures for such models used in this paper are based on the computationally efficient Fisher scoring algorithm (Longford, 1987) implemented in the software VARCL (Longford, 1986). It provides estimates of regression parameters and (co-) variances, together with standard error for them, and the value of the log-likelihood, which permits formal likelihood ratio hypothesis testing.

The Sample

The IEA Second International Mathematics Study (SIMS) sample comprised of 99 mathematics teachers and their 4030 eighth-grade students and was derived from a two-stage, stratified random sample of classrooms. The thirteen primary sampling units were the twelve national educational regions of Thailand plus the capital, Bangkok. Within each region, a random sample of lower-secondary schools was selected. At the second stage, a random sample of one class per school was selected from a list of all eighth grade mathematics classes within the school. The resulting sample represented a 1% sample of eighth grade mathematics classrooms within each region. This region, of course, does not distinguish between the school and classroom levels, and so only inference about the aggregate of these effects is possible.

At both the beginning and end of the school year, students were administered a mathematics test covering five curriculum content areas (arithmetic, algebra, geometry, statistics and measurement). Teachers completed several instruments at the posttest, including a background questionnaire and a general classroom process questionnaire. Teachers provided information about teaching practices and characteristics of their randomly selected "target" class. Data about the school was provided by a school administrator. In the following sections, a description of each of the variables analyzed in this paper is provided (see Lockheed, Vail and Fuller, 1987, for a more extended discussion); acronyms for the variables are given in parentheses. For easier orientation, the acronyms for pupil-level variables are given in capital letters and for group-level (region/school/classroom) variables in lower case letters. This will be clear from Tables 1 and 2, which provide definitions and summary statistics for all variables.

Measures

Mathematics achievement. The IEA developed five mathematics tests for use in SIMS. One of the tests was a forty-item instrument called the core test. The remaining four tests were thirty-five item instruments called rotated forms and designated A through D. The five test instruments contained roughly equal proportions of items from each of the five curriculum content areas, except that the core test contained no statistics items. For purposes of this analysis we regard the instruments as parallel forms with respect to mathematics content.

The IEA longitudinal design called for students to be administered both the core form and one rotated form chosen at random at both pretest and posttest. In Thailand, students were pretested using the core test and one rotated form. At posttest, students again took the core test and one rotated form, but were prevented from repeating the rotated form taken at pretest. Approximately equal numbers of students took each of the rotated forms in both administrations.

One goal of this analysis was to predict posttest achievement as a function of pretest performance and of other determinants. Since students took the core form twice, the core form posttest score reflects, to some degree, familiarity with the core test items. Instead of using the core test, therefore, we analyzed scores obtained from the rotated forms, after they were equated to adjust for differences in test length and

Table 1: Variable Names, Descriptions and Means (Proportions) of Student-Level Variables for Three Data Sets

Variable Name		Means/Proportions Data Set 1	Data Set 2	Data Set 3
Sample				
Students		2076	2804	3025
Classroom		60	80	86
Student-Level Variables				
XROT	Pretest mathematics achievement score	9.15	8.83	8.83
XAGE	Age in months	170.94	171.05	171.09
XSEX	Student sex (0 = female; 1 = male)	.53	.53	.53
YFOCCI	Father's occupational status:			
	Unskilled or semi-skilled worker	.15	.15	.15
	Skilled Worker	.44	.45	.46
	Clerical or sales worker	.26	.26	.25
	Professional or managerial worker	.15	.15	.14
YMEDUC	Mother's educational attainment			
	Very little or no schooling	.26	.26	.26
	Primary school	.58	.58	.58
	Secondary school	.09	.09	.09
	College, university or some form of tertiary	.07	.07	.06
HCALC	Calculator at home (0 = no; 1 = yes)	.31	-	-
YHLANG	Use language of instruction at home (0 = no; 1 = yes)	.49	-	-
YMOREED	Educational expectation			
	Less than two years	.08	.08	.08
	Two to four years	.30	.31	.30
	Five to seven years	.41	.41	.41
	Eight or more years	.22	.20	.21
YPARENC	Parental encouragement (1 = high)	2.12	2.10	2.09
YPERCEV	Perceived mathematics ability (1 = high)	4.05	4.05	4.05
YFUTURE	Perceived future importance of mathematics (1 = low)	2.06	2.05	2.06
YDESIRE	Motivation to succeed in mathematics (1 = low)	5.47	5.47	5.47

difficulty. In this analysis, we used equated rotated form formula scores for both pretest (XROT) and posttest (YROT) measures of student mathematics achievement.[3]

Student background characteristics. Basic background information about each student included his or her sex (XSEX), age in months (XAGE), highest maternal education (YMEDUC), paternal occupational status (FOCCI), home language (YHLANG) and home use of a four-function calculator (YHCALC). Paternal

Table 2: Variable Names, Descriptions and Means (Proportions) of Group-Level Variables for Three Data Sets

Variable Name		Means/Proportions Data Set 1	Data Set 2	Data Set 3
Sample				
Students		2076	2804	3025
Classroom		60	80	86
Student-Level Variables				
SPCI81	District per capita income (in 1000 bahts)	12.94	12.97	-
SENROLT	Number of students in school (in 1000)	1.27	1.44	1.41
SSTEAM	Ability groupings for instruction (0=no; 1=yes)	.46	.47	-
SDAYSYR	Days in school year	195.04	-	-
SPUTEAR	Pupil-teacher ratio in school	14.86	15.81	15.93
SQUALMT	% of teachers in school qualified to teach math	.57	.62	.62
TECMATH	Semesters of post-secondary mathematics	3.95	-	-
TSEX	Teacher sex (0 = female, 1 = male)	.33	.37	-
TAGE	Teacher age in years	29.04	-	-
TEXPTCH	Years of teaching experience	7.25	-	-
TNSTUDS	Years of students in target class	43.61	42.61	-
TMTHSUB	Math curriculum (0=remedial or normal, 1=enriched)	.22	.20	.18
TXTBK	Frequent use of textbook (0=no; 1=yes)	.55	.56	.58
CEFEED	Frequent individual feedback	2.15	-	-
TWORKBK	Use of published workbooks (0=no; 1=yes)	.85	.83	.81
TVISMAT	Use of commercial visual materials (0=no; 1=yes)	.34	.40	-
TADMIN1	Weekly minutes spent in routine administration	26.84	-	-
TORDER1	Weekly minutes spent in maintaining class order	19.40	20.27	20.33
TSEAT1	Weekly minutes students spent at seat or blackboard	53.76	54.57	-

occupation was classified into four categories; (a) unskilled or semi-skilled worker, (b) skilled worker, (c) clerical or sales worker, and (d) professional or managerial worker. Highest maternal education was also classified into four categories: (a) very little or no schooling, (b) primary school, (c) secondary school, and (c) college, university or some form of tertiary education.

Student attitudes and perceptions. Five indices of student attitudes and perceptions were also included. Student educational expectations (YMOREED) were measured by a single item that asked about the number of years of full-time education the student expected to complete after the current academic year. The following categories were defined: (a) less than two years, (b) two to four years, (c) five to seven years, and (d) eight or more years. Parental encouragement (YPARENC) was measured by a four-item index composed of responses on a Likert-type scale in which students described their parent's interest in, and encouragement for, mathematics achievement. For example, for the item "My parents encourage me to learn as much mathematics as possible"; response alternatives ranged from "exactly like" the student's parents (= 1) to "Not at all like" the student's parent (= 5). The four items comprised a single factor, with principal component factor loadings ranging from .72 to .83 and communality of 2.43. A low score represented greater parental support. Perceived mathematics ability (YPERCEV), perceived usefulness of mathematics (YFUTURE), and motivation toward mathematics achievement (YDESIRE) were all developed from a factor analysis of the student attitude survey, which contained Likert-type items having response alternatives ranging from "strongly disagree" (= 1) to "strongly agree" (=5). Factors were initially identified through VARIMAX factor analyses, and then confirmed through principal component analyses, from which factor scores were constructed. For YPERCEV, a low value represents a positive attitude; for YFUTURE and YDESIRE a high value represents a positive attitude.

School characteristics. Data on six school characteristics are analyzed in this paper: (a) school size, as indicated by the total number of students enrolled in the school (SENROLT), (b) presence of ability grouping (SSTREAM), (c) length of the school year indays (SDAYSYR), (d) student teacher ratio in the school (SPUTEAR), (e) percentage of the teaching staff qualified to teach mathematics (SQUALMT), and district-level per capita income in 1981 (SPCI81).

Teacher characteristics. Four teacher characteristics are analyzed: (a) sex of the teacher (TSEX), (b) remedial or typical versus enriched mathematics subject matter (TMTHSUB), and (c) whether or not the teacher used textbooks frequently in the class (TXTBOOK).

Teaching practices. Six variables referring to teaching practices are considered: (a) providing feedback to students (a composite index of five elements of teaching practice: commenting on student work, reviewing tests, correcting false statements, praising correct statements, and giving individual feedback) (CEFEED); number of minutes per week the teacher spent on (b) routine administration (TADMIN1), (c) maintaining class order (TORDER1), (d) monitoring assigned seatwork (TSEAT1); (e) using commercially produced visual materials (TVISMAT), and (f) using workbooks (TWORKBK). In summary, the data contain information on 32 variables about 4030 pupils from 99 schools. Of the 32 variables, 13 are student characteristics, 5 variables refer to the school, 4 to the teacher, 9 variables are defined for the classroom, and one variable is a characteristic of the district (catchment area). The distinction between variables defined for pupils and for classrooms/teachers/schools (henceforth groups, since they are confounded in the design) is important because they play different roles in explanation of variation. Also, it should be noted that the complete data set consists of 13*4030 + 19*99 = 54,271 units of data, although conventionally it would be conceived, and stored on a computer, as a data set with 32*4030 = 128,960 units of data. The data contain relatively more information about the groups (19 variables for 99 units) than for the pupils (13 variables for 4030 units). Arguably, group-level variables are also more reliable, because they refer to school or teacher records, and are responses from adult professionals, whereas the responses of pupils are subject to test-performance variation, recall of family circumstances and arrangements, variable interpretation of the questionnaire items, and so on. Also pupil-level variables, e.g., SES or XROT, have a large group level component of variation; groups vary a great deal in their composition (means, standard deviation, etc.) of these variables. Hence, not only the 19 group-level variables, but also to some extent the 13 pupil-level variables potentially explain group-level variation among the 99 groups, whereas only the 13 pupil-level variables can explain some of the pupil-level variation of the outcome scores of 4030 pupils.

RESULTS

The response rate for the 13 pupil-level variables is between 93-100 percent. There is no obvious pattern of missingness among the pupils; complete pupil-level records are available for 3466 individuals (86%). The group-level data are available for between 78-99 schools, but only 60 schools have complete records, and within these schools only 2076 pupils also have complete pupil-level data (51.5%).

Our intention is to carry out a multiple regression analysis of the data, and seek a linear prediction formula for the posttest scale score (YROT) in terms of the pretest scale score (XROT) and a suitable subset of the 30 other (explanatory) variables. For a model which involves a given set of variables we would use the data on all pupils and schools, for whom all the responses on the variables in the set are available (listwise deletion). Thus for a smaller, more parsimonious, set of variables we have a larger sample of pupils and schools.

Our general strategy in this modelling approach is as follow: we start with the data set obtained by listwise deletion with respect to all variables (2076 pupils in 60 schools), fit regression models to this data set, apply a conservative criterion (to be specified below) to exclude variables from the obtained regression formula, thus constructing a restricted set of explanatory variables. For this restricted set of variables (including the outcome YROT) we apply listwise deletion, which leads to a larger sample of pupils and schools. For this new data set we again fit regression models, simplify the regression formula, if possible, and continue on until no further reduction of the set of variables, and extension of the data set obtained by listwise deletion , is possible.

Usually it cannot be assumed that the unavailable data are missing at random, i.e., the distribution of a variable among the pupils from whom we obtain valid responses is similar to the distribution among the pupils whose responses are not available (missing). In educational surveys, typically, higher ability pupils, those with higher social status, etc., tend to have higher response rates, implying bias in estimates of certain population means, as well as in regression coefficients obtained from simple regression. Missingness at random is an unnecessarily stringent criterion for ensuring that omission of the subjects with missing data has no effect on the results of a regression analysis. It is sufficient to have conditional randomness, given the explanatory variables. It means that for any combination of explanatory variables the distribution of the outcome among the pupils in the sample is identical to those excluded from the sample by the listwise deletion procedure. Intuitively, such an assumption becomes less stringent the more explanatory (conditioning) variables are used. On the other hand, a larger set of explanatory variables implies a larger proportion of subjects whose data are not used in the analysis.

An indication of the extent to which the criterion of conditional randomness is relevant can be deduced from comparisons of model fits for two different samples: the maximal sample obtained by listwise deletion with respect to the set of explanatory variables used in the considered model, and the sample obtained by listwise deletion with respect to a more extensive, or complete, set of explanatory variables. In a few of such comparisons, reported below, we found a close agreement in several pairs of such analyses.

Variance Component Models

The hierarchical structure of the data, with pupils nested within groups, requires a form of regression analysis which takes into account the two separate sources of variation. Separation of the variation due to pupils and due to schools/classrooms is also of substantive interest, because the latter is a measure of the size of unexplained differences among the schools/classrooms.

Relevance of variance component methods for analysis of data with hierarchies has been established by Goldstein (1986), Raudenbush and Bryk (1986 and Aitkin and Longford (1986); they address the previously-mentioned problems with the use of the ordinary regression methods when the assumption of independence of the observations is not satisfied.

Variance component models compared with OLS. Variance component methods involve the explicit modelling of the student and group variation, and afford flexibility of modelling of the group variation, which cannot be allowed for in ordinary regression. The specification of a variance component model is necessarily more complex than for the ordinary regression. In standard situations, first the list of the regression variables involved in explanation of the outcome for a typical (average) group has to be declared, and then a sublist of this list should be declared, which contains the variables for which the within-group relationships vary from group to group. The full list of variables, referred to as the FIXED PART, is analogous to the list of the explanatory variables in ordinary regression. The sublist (RANDOM PART) may contain only pupil-level variables, which are not constant within all the groups, because within-group regression coefficients on group-level variables cannot be identified.

Variance component models involve two kinds of parameters. The fixed effects parameters refer to the regression relationship for the average group. Their interpretation is analogous to the regression parameters in the ordinary regression. The random effects parameters are variances and covariances that describe the between-group variation in the regression relationship. Of prime interest are the sizes of the variances. Zero variance of a regression coefficient corresponds to constant relationship across the groups. In order to obtain information about the variation we require, in general, a substantially larger number of pupils and groups than for the regression parameters. We can therefore expect to find a small random part, containing only a few variables, as a sufficient description of the variation, whereas the fixed part may contain most of the available explanatory variables.

One important aspect of the separation of the two sources of variation is in distinguishing between pupil and pupil-level variation. This comes out very clearly in the following examples: it turns out that we have abundant group-level information, i.e., a good description of the between-group variation, but a much larger proportion of the student-level variation remains unexplained.

To fix ideas, we consider first a specific model

$$y_{ij} = e_k\, x_{ij,k}\, b_k + d_j + e_{ij}$$

where the indices $i = 1, ..., n_j = 1, .., N_2, k = 1, ..., K$, represent the pupils, groups, and the variables, respectively. The b's are the regression parameters, and the d's and e's are the group- and pupil- level random effects, and are assumed to be independent random samples from the normal distribution with zero means and variances s^2 and t^2. In analogy with the ordinary regression we can define the R^2 as the proportion of variation explained as

$$R^2 = 1 - (s^2 + t^2)/(s^2_{raw} + t^2_{raw}),$$

where the subscript "raw" refers to the variance estimates in the "empty" variance component model

$$Y_{ij} = m + d_j + e_{ij}$$

It is advantageous, however, to define two separate R^2's which refer to the two levels of the hierarchy:

$$R_p^2 = (1 - s^2)/s^2_{raw}$$

$$R_g^2 = (1 - t^2)/t^2_{raw}$$

for pupils and groups, respectively.

<u>Example 1: Ordinary regression</u>. In the present analysis, for a data set obtained by listwise deletion with respect to a set of variables considered below (3136

pupils in 88 schools) we have for the simple regression of posttest (YROT) on pretest (XROT):

$$s^2_{raw} = 82.80$$

$$E\{YROT\} = 4.892 + \underset{(.015)}{.818}\, XROT$$

$$s^2 = 42.56,$$

and so $R^2 = 1 - s^2/s^2_{raw} = .486$.

The standard errors for the regression estimates will be given throughout the paper in parentheses in the line below the regression parameters. For example, .015 above is the standard error for the regression coefficient on XROT, .818. The corresponding t-ratio is .818/.015 = 54.5. In this model, identification of pupils within schools is completely ignored, and the pupils are assured to be a randomly drawn sample from the population of all pupils in a given grade in the country. A pupil with a given pretest score X is expected to score 4.892 + .818X on the posttest administration. This pupil would be, however, likely to have a score quite substantially different from this prediction, because the variance of all the pupils with a given pretest score is $s^2 = 42.56$ (standard deviation — $\sqrt{42.56} = 6.5$). The prediction is still a marked improvement if we only used the overall mean of the YROT scores, 12.2, as a prediction for the pupil. Then the standard deviation would be 9.1 $\sqrt{82.80}$.

Since in future text it will be clear from the context whether the parameter or its estimate is meant, the "^" notation will be abandoned.

<u>Example 2: (Simple) Variance component model:</u>

$$Y_{ij} = m + d_j + e_{ij}$$
$$s^2_{raw} = 55.56$$
$$t^2_{raw} = 25.65$$

The variation of posttest scores has a substantial group-level component; the variance component ratio is r = 25.65/81.21 = .316. The variance component regression model is given as:

$$E\,[YROT] = 5.841 + \underset{(.018)}{.699}\,XROT$$

$$s^2 = 38.55$$
$$t^2 = 4.78,$$

and so we have $R^2 = 1 - 43.44/81.21 = .466$, and

$$R_p{}^2 = 1 - 38.55/55.56 = .306$$
$$R_g{}^2 = 1 - 4.78/25.65 = .814.$$

Thus, if we make allowance for the within school homogeneity of the posttest scores, we obtain a prediction formula for the posttest score (Y = 5.841 + .699X) that is substantially different from the OLS regression obtained in Example 1. Note also by how much the school-level variation has been reduced. Table 3 presents the comparison between the simple OLS and simple variance component models. Clearly the latter extension of the R^2 for variance components is more informative. The pretest score XROT is a powerful predictor of the posttest score YROT. But whereas it explains more than 80% of the variation among the groups, the proportion of the pupil-level variation explained is only 30%. The school-level variation in the outcome scores reflects the pretest score to a great extent. Some of the remaining within-group variation may be explained by the other explanatory variables, but they are not likely to have as dominant an effect as the pretest score. The variation associated with the testing and scoring procedure, which could be demonstrated in an experiment with repeated administration of the test, use of alternate fores, etc., will remain as a component of the pupil-level variation. Thus, whereas group-level variation can potentially be reduced to 0, pupil-level variation has a component that cannot be explained by any explanatory variables. In ideal circumstances (and in our case, almost) we can explain completely why 'how schools

Table 3: Comparison of OLS and VCS Models of Grade 8 Mathematics Post-Test Predicted From Pretest, Thailand 1981-82

Model	Method	
	OLS	VCS
Empty model		
σ^2_{raw}	82.80	55.56
σ^2_{raw}	-	25.65
Regression model		
Intercept	4.892	5.841
Coefficient	0.818	0.699
St. error coeff.	0.015	0.018
σ^2	42.56	38.55
τ^2	-	4.78
R^2	0.486	-
R_p^2	-	0.306
R_g^2	-	0.814

vary; the variance of schools in the later models is very small. But pupil-level variation cannot be completely explained; there will always be the unexplained (and in our case unidentifiable) <u>within-pupil</u> variation. Since every pupil provides only one outcome score, the within-pupil and within group variation cannot be separated.

The raw variance component ratio is .316, but for the model with the pretest score the ratio drops to .110. If pretest score is ignored, groups appear to have substantial differences. But schools appear to be much more similar (homogeneous) once we take account of the pretest scores, i.e., they are much more similar in the way they "convert" initial ability into outcome.

If a group-level explanatory variable were added to the regression model, it would result in a reduction of only the group-level variance, which has already been substantially reduced. Therefore there is a limited scope for important group-level explanatory variables. By comparison, among the pupil-level variables there may be ones that explain a great deal of the remaining pupil-level variation.

Inclusion of a pupil-level variable in the regression model will cause a reduction of both the pupil- and group-level variances. The relative sizes of the reductions of the two variances will depend on how the variation of the explanatory variable decomposes into between- and within-group variance. Hence the potentially most important pupil-level explanatory variables are those with little between-group variation.

Example 3: Variable slopes model. The variance component model discussed above can be further generalized to the model which allows variable slopes on the pretest:

$$Y_{ij} = b_0 + b_1 x_{ij} + d_{oj} + d_{1j}(x_{ij} - x) + e_{ij},$$

where (d_{oj}, d_{1j}) form a random sample from $N(O, S_d)$ and e's are i.i.d. $N(0, \sigma^2)$. The maximum likelihood estimates for this model are:

$b_0 = 5.832$

$b_1 = .687\ (.019)$

$\sigma^2 = 38.367$

$$S_d = \begin{pmatrix} 4.947 & \\ .0805 & .00416 \end{pmatrix}$$

The software VARCL used for maximum likelihood estimation in variance component models estimates the square root of the variances in S_d, and produces standard errors for these estimates:

$S_{d,11} = 2.224\ (.202)$

$S_{d,22} = .0645\ (.0338)$

$S_{d,12} = .0805\ (.0311)$

The value of the deviance (-2 log-likelihood) is 20,496.3. Using the conventional t-ratio we conclude that the slope-variance $S_{d,22}$ is not significantly different from 0, and so we can adopt the simple variance component model. More formally we can use the likelihood ratio test for comparison of the two variance component models. The deviance for the simple model is 20499.9, 3.6 higher than for the model with variable slope. The simpler model is obtained from the latter model by constraining to zero the slope variance $S_{d,22}$ and the slope-by-intercept covariance $S_{d,12}$. The fitted correlation of the slope and intercept is .56; the variance matrix S_d is non-singular. Constraints on the two parameters (degree of freedom) have led to an increase of the deviance of only 3.6 (to be compared with the chi-square tables of critical values for 2 d.f.), and hence we can declare that we have found insufficient evidence for variable slope of the posttest on pretest among the schools. The differences among the schools, described by the variance t^2 in the simple variance component model, are substantial, and statistically significant; the formal likelihood ratio test for the hypothesis that $t^2 > 0$ is obtained by comparison of the deviance of the ordinary regression and the simple variance component models. The ordinary regression deviance (-2 log-likelihood, not the same as the residual sum of squares!) is equal to 20662.6, 162.6 higher than the deviance for the simple variance component model (chi-square with 1 degree of freedom). Also the t-ratio for t^2 is larger.

Making inference about variable relationships is of substantive importance in school effectiveness studies. Schools are expected to vary in their performance, after accounting for differences in the initial ability of the pupils, but other more complex patterns of between-school variation may arise: Schools may be relatively more successful in teaching children with certain background characteristics, they may either exaggerate, or reduce differences among the pupils at enrollment.

Variable relationships are intimately connected with variance heterogeneity. For illustration, we consider the variable slope model discussed above. The fitted variance of an observation is

$$38.367 + 4.947 + 2^*(XROT - 8.912)^*.08054 + (XROT - 8.912)^2 {}^*.00416;$$

it is a quadratic function of the pretest. The minimal variance occurs for XROT*= 8.912 - .0805/.0042 = 10.45, and is equal to 41.75. Only two pupils in the whole sample have scores lower than XROT*. Larger values of the explanatory variable XROT are associated with larger variance. For XROT = 9 (near the mean) the fitted variance is 43.33, and for XROT = 30 (near the sample maximum) the fitted variance is 48.56. It would appear that for low-ability pupils the choice of the school they attend is slightly less important than for high-ability pupils. We have to bear in mind, though, that we are dealing with an observational study, not with an experiment, and in reality pupils, or their parents, do not exercise completely free choice over the school. Thus a causal statement, or a prediction about a future manipulative procedure, can be made only under the condition that all the other circumstances in the educational system remain intact. This is usually a very unrealistic assumption.,

Comparison of models. The comparison of the regression relationship (fixed effects) is instructive. We have

1. Ordinary regression:

   E [YROT] = 4.892 = .818*XROT
   (.015)

2. Simple variance component model

   E [YROT =5.841 + .699*YROT
   (.017)

3. Variable slopes

   E [YROT] = 5.832 + .687*XROT
   (.019)

The estimate of the regression coefficient on XROT in ordinary regression is substantially different from the estimates in the two variance component models. Ignoring the hierarchical structure of the data would lead to different conclusions, say, for prediction of posttest *YROT) from pretest (XROT). In other words, whereas the OLS estimate could be interpreted to mean that each point on the pretest is worth .82 points on the posttest, the VCS estimate more accurately places this value at .69 points.

Multiple Regression Models

The purpose of this section is to obtain the most parsimonious simple variance component model of grade 8 mathematics learning in Thailand, given the available data.

We proceed as follows. First we fit the simple variance component model using the largest data set obtainable by listwise deletion with respect to a given set of variables. Second, we apply an exclusion criterion, defined below, to eliminate variables from the model, creating a new model, and then we fit this new model on the same data. These three steps are repeated, with listwise deletion with respect to the restricted set of variables, until no more variables can be eliminated.

Regression with all the variables. We begin with fitting simple variance component models (VCS), i.e., models involving no variable slopes, to the data set obtained by listwise deletion with respect to all the available variables. This data set contains 2076 pupils in 60 schools.

The ordinary regression fit (OLS) of the posttest on pretest is

$$E[YROT] = 4.882 + \underset{(.017)}{.817}{*}XROT, \qquad s2 = 42.20,$$

which is in close agreement with the OLS fit reported above for a larger data set (3136 pupils in 88 schools). The corresponding simple variance component model fit is:

$$E[YROT] = 5.670 + \underset{(.020)}{.720}{*}XROT$$

$$s^2 = 38.79$$
$$t^2 = 4.02$$

Compared to the larger data set, we find some discrepancies: the fitted regression slope for the smaller data set is higher (.720 vs. .699), and the group-level variance is smaller (4.02 vs. 4.78). Variation of the slope on XROT is not significant in either sample, but it is two-and-a-half times as great as the larger data set (.00416) than in the smaller one (.00166). It appears that the 28 schools added to the data are more likely to have lower regression slopes, and contain proportionately more extreme schools (very "good" or very "bad"), because the larger sample has larger group-level variance t^2. We emphasize that all these differences may arise purely by chance, rather than as a result of non-random missingness of data, but they can have a substantial effect on the inferences drawn.

The OLS and VCS model estimates for the 2076/60 data using all the explanatory variables are given in Display 1. The dominant explanatory power of the pretest score XROT is obvious, judging not only by the t-ratio for its regression coefficient (32.38 for OLS and 30.80 for VCS), but also by the comparison of the variance component estimates across models. The raw variance component estimates are:

$$s_{raw}^2 = 57.30$$
$$t_{raw}^2 = 22.83$$

The pretest score XROT on its own leads to reduction of these variances to 38.79 ($R_p^2 = 32\%$) and 4.02 ($R_g^2 = 86\%$), but the other 30 variables reduce the pupil-level variance only marginally (to 36.8, R_p^2 - 36%). The group-level variance is almost saturated (1.32, $R_g^2 = 95.5\%$). It appears that we have abundant information about the groups, but we are less successful in expl ation, or suitable description, of pupil-level variation.

The relatively large number of group-level variables raises the concern about multicollinearity, i.e., competing alternative descriptions of the data. To deal with this problem we apply a conservative criterion for exclusion of explanatory variables from our models. We regard a variable as not "important" for the fixed part of the VCS model, if the t-ratio of its regression coefficient is smaller than 0.9 at the first stage of model reduction and 1.0 thereafter. In the first round of simplifying the model, we use the 0.9 criterion to exclude two pupil-level variables (HCALC and YHLANG) and six group-level variables (SDAYSYR, TECMATH, TAGE, TEXPTCH, CEFEED, and TADMIN1) from the full list of 31 variables.

Second model. Next we estimate a VCS model fit with this shorter list of 23 variables. The results are shown in Display 2. Exclusion of these variables (8 degrees of freedom) has virtually no effect on the retained regression parameters and their standard errors (compare Displays 1 and 2; the exception is TVISMAT, which now fails to meet the inclusion criterion), and the increase in the variance components is only marginal in particular for the group-level variance. The difference in deviances is 3.3 (c_8^2).

Then we obtain the largest data set obtainable by listwise deletion with respect to the retained variables; this yields data for 2804 pupils in 80 schools. We then compute the variance component analysis for th data set; results are given in

Display 1: OLS and VCS Model Estimates for 2076 Students and 60 Classrooms/Schools Using All 31 Explanatory Variables, Thailand 1981-82

Variable	OLS Estimate	OLS St. Error	VCS Estimate	VCS St. Error
GRAND MEAN	18.603	-	19.717	-
XROT	.680	.021	.647	.021
XAGE	-.080	.016	-.077	.016
XSEX	.732	.301	.969	.319
YFOCCI	.174	.431	.033	.434
	-.631	.462	-.646	.460
	-.178	.541	-.239	.542
YMEDUC	.021	.327	-.039	.325
	-.129	.562	-.157	.556
	-.686	.661	-.899	.663
HCALC	-.120	.310	-.217	.309
YHLANG	.203	.315	.012	.341
YMOREED	1.087	.546	1.074	.541
	1.570	.545	1.537	.541
	1.638	.593	1.610	.589
YPAREBC	.225	.137	.249	.136
YPERCEV	-.980	.160	-1.020	.161
YFUTURE	.574	.168	.526	.167
YDESIRE	.277	.236	.228	.233
SPCI81	.061	.042	.073	.060
SENROLT	.422	.263	.417	.386
SSTEAM	-.426	.358	-.509	.512
SDAYSYR	-.006	.020	-.010	.029
SPUTEAR	-.152	.051	-.170	.075
SQUALMT	1.023	.342	.1.029	.494
TECMATH	-.035	.037	-.044	.053
TSEX	-.580	.336	-.619	.481
TAGE	.009	.032	-.001	.046
TEXPTCH	.014	.043	.038	.064
TNSTUDS	.035	.018	.039	.025
TMTHSUB	1.725	.432	1.941	.628
TXTBOOK	1.602	.338	1.650	490
CEFEED	.148	.203	.209	.290
TWORKBK	-1.104	.218	-1.124	.314
TVISMAT	.380	.331	.461	.480
TADMIN1	-.003	.004	-.003	.006
TORDER1	-.037	.012	-.039	.016
TSEAT1	.011	.005	.011	.007
Variance	38.031	6.167	-	-
Pupil-level Variance	-		36.809	-
Pupil-level Sigma	-		6.067	-
Group-level Variance	-		1.317	-
Group-level Sigma	-		1.148	0.192
Deviance			13424.947	-

Display 3. We see that the regression coefficients for the pupil-level variables are stable across the data sets (compare with Displays 1 and 2), but for the group-level variables there are substantial discrepancies. There are two separate, but possibly complementary, explanations for these discrepancies: multicollinearity and non-random missingness of data. Multicollinearity would cause the regression estimates to be sensitive to changes in the data, in our case to inclusion of over 700 new observations. As an alternative, the discrepancies could arise as a result of the non-random missingness in our data, i.e., if the two data sets have genuinely different regression characteristics. A suitable indication, though not a fool-proof check, for the latter possibility is obtained by fitting of models with identical specifications for the different "working" data sets. We have fitted the reduced second model (Display 2) to the larger data set (Display 3), and although different values of the group-level regression coefficients were obtained, it turns out that the reduced list of variables also provides an adequate description for the data (as judged by the likelihood ratio criterion). The pupil-level regression coefficients differ only marginally.

We conclude, therefore, that multicollinearity is the more likely cause of the discrepancies in the estimates; we have too many group-level variables, and so the parameter estimates are subject to large fluctuation with small changes in the data. The explanatory variables provide sufficient conditioning for the outcome data to be missing at random given the available explanatory variables.

Display 2: OLLS and VCS Model Estimates for 2076 Students and 60 Classroom/Schools Using 23 Explanatory Variables, Thailand 1981-82

Variable	OLS Estimate	OLS St. Error	VCS Estimate	VCS St. Error
GRAND MEAN	18.118	-	18.370	-
XROT	.685	.020	.650	.021
XAGE	-.080	.016	-.076	.016
XSEX	.723	.299	.958	.318
XFOCCI	.118	.426	.033	.432
	-.621	.457	-.651	.457
	-.139	.538	-.212	.541
YMEDUC	.037	.326	-.028	.325
	-.068	.559	-.115	.555
	-.604	.656	-.855	.660
YMOREED	1.115	.545	1.083	.540
	1.568	.543	1.521	.540
	1.666	.591	1.609	.589
YPARENC	.238	.137	.255	.135
YPERCEV	-.970	.160	-1.010	.161
YFUTURE	.570	.168	.526	.167
YDESIRE	.287	.235	.234	.233
SPCE81	.050	.038	.058	.056
SENROLT	.509	.251	.540	.373
SSTEAM	-.441	.324	-.503	.472
SPUTEAR	-.178	.046	-.198	.068
SQUALMT	1.062	.327	1.090	.430
TSEX	-.518	.314	-.536	.460
TNSTUDS	.036	.017	.038	.025
TMTHSUB	1.802	.409	2.094	.604
TXTBOOK	1.649	.315	1.673	.463
TWORKBK	-.1028	.204	-1.039	.300
TVISMAT	.368	.322	.393	473
TORDER1	-.040	.010	-.043	.014
TSEAT1	.010	.005	.011	.007
Variance	38.108	6.173	-	-
Pupil-level Variance	-		36.855	-
Pupil-level Signa	-		6,071	-
Group-level Variance	-		1.351	-
Group-level Sigma	-		1.162	.191
Deviance	-		13,428.295	-

Display 3: OLS and VCS Model Estimates for 2804 Students and 80 Classrooms/Schools Using 23 Explanatory Variables, Thailand 1981-82

Variable	OLS		VCS	
	Estimate	St. Error	Estimate	St. Error
GRAND MEAN	17.659	-	17.314	-
XROT	.699	.017	.634	.019
XAGE	- 079	.014	-.073	.014
XSEX	.746	.251	1.103	.271
YFOCCI	.197	.363	.101	.367
	-.403	.389	-.458	.386
	.089	.458	.085	.458
YMEDUC	.306	.279	.293	.276
	.088	.465	.142	.458
	-.018	.567	-.309	.566
YMOREED	.861	.476	.786	.467
	1.086	.475	1.015	.468
	1.617	.519	1.542	.512
YPARENC	.388	.118	.375	.116
YPERCEV	-.1083	.137	-1.131	.136
YFUTURE	.576	.142	.533	.141
YDESIRE	.493	.201	.439	.198
SPCI81	-.029	.033	0.035	.057
SENROLT	.437	.187	.481	.331
SSTEAM	-.417	.275	-.422	.473
SPUTEAR	-.095	.032	-.110	.058
SQUALMT	.0698	.246	.784	.429
TSEX	-.038	.266	.014	.463
TNSTUDS	.012	.014	.020	.023
TMTHSUB	1.836	.344	2.398	.593
TXTBOOK	.948	.266	.978	.461
TWORKBK	-.0.500	.167	.-.499	.291
TVISMAT	.353	.269	.363	.468
TORDER1	-.024	.008	-.027	.013
TSEAT1	.005	.004	.006	.006
Variance	37.949	6.160	-	-
Pupil-level Variance	-		35.868	-
Pupil-level Sigma	-		5.989	-
Group-level Variance	-		2.285	-
Group-level Sigma	-		1.512	0.174
Deviance	-		18088.395	-

According to our exclusion criterion (t-ratio < 1) we now delete from the fixed part of the model the following six group-level variables: SPCI81, SSTREAM, TSEX, TNSTUDS, TVISMAT, and TSEAT1.

Third model. As before, we estimate this model with both smaller and larger data sets. For the former, OLS and VCS model estimates for this reduced list of variables are given in Display 4; the same schools and pupils are involved as for Display 3. For the latter, 3025 students in 86 schools, we fit the reduced model (17 variables). The results are given in Display 5. Again, the difference in deviances (3.5, c_6^2) is small. The effects of non-random missingness can be checked by comparison of the estimates in Displays 4 and 5. Applying our exclusion criterion to the variables in this model, we find that no further reduction of the list of explanatory variables is now possible.

We note that, owing to the relatively small number of schools, the appropriate conclusion about the 14 group-level variables is that we "have found insufficient evidence" of a systematic effect of these variables, rather than "our analysis disproves their effects". Also, a different modelling scheme could lead to a different "minimal" set of important explanatory variables. Because of collinearity, there may be a set of alternative regression formulas that give a model fit which is not substantially inferior to the one given in Display 5, in terms of the deviances. A summary of the results of these analyses is provided in Table 4.

Display 4: OLS and VCS Model Estimates for 2804 Students and 80 Classrooms/Schools Using 17 Explanatory Variables, Thailand 1981-82

Variable	OLS Estimate	OLS St. Error	VCS Estimate	VCS St. Error
GRAND MEAN	17.321	-	17.694	-
XROT	.704	.017	.635	.018
SAGE	-.077	.014	-.073	.014
XSEX	.676	.247	1.086	.270
YFOCCI	.181	.357	.085	.365
	-.419	.387	-.465	.385
	.105	.455	.082	.457
YMEDUC	.293	.280	.288	.276
	.112	.465	.154	.458
	.014	.563	0.297	.564
YMOREED	.869	.476	.786	.467
	1.128	.476	1.027	.468
	1.666	.520	1.560	.512
YPARENC	.393	.117	.377	.116
YPERCEV	-.1.076	.137	-1.130	.136
YFUTURE	.592	.142	.537	.141
YDESIRE	.477	.201	.431	.197
SENROLT	.285	.164	.367	.289
SPUTEAR	-.074	.030	-.094	.054
SQUALMT	.808	.239	.880	.427
TMTHSUB	1.950	.329	2.562	.576
TXTBOOK	.948	.259	.946	.458
TWORKBK	-.433	.160	-.402	.284
TORDER1	-.022	.006	-.024	.010
Variance	38.065	6.170	-	-
Pupil-level Sigma	-		35.871	-
Pupil-level Variance	-		5.989	-
Group-level Variance	-		2.429	-
Group-level Sigma	-		1.558	0.176
Deviance	-		18091.983	-

Display 5: OLS and VCS Model Estimates for 3025 Students and 86 Classrooms/Schools Using 17 Explanatory Variables, Thailand, 1981-82

Variable	OLS Estimate	OLS St. Error	VCS Estimate	VCS St. Error
GRAND MEAN	17.238	-	17.536	-
XROT	.695	.017	.629	.018
XAGE	-.075	.014	-.071	.014
XSEX	.658	.238	1.053	.260
YFOCCI	.152	.343	-.435	.373
	-.415	.373	-.435	.373
	.115	.443	.123	.446
YMEDUC	.371	.269	.343	.265
	.056	.449	.073	.442
	.066	.554	-.259	.555
YMOREED	.854	.461	.755	.453
	1.195	.459	1.064	.452
	1.703	.500	1.532	.494
YPARENC	.361	.113	.347	.112
YPERCEV	-1.140	.132	-1.191	.132
YFUTURE	.614	.137	.543	.136
YDESIRE	.484	.194	.459	.190
SENROLT	.271	.160	.350	.279
SPUTEAR	-.076	.029	-.094	.052
SQUALMT	.847	.232	.903	.410
TMTHSUB	1.968	.327	2.546	.566
TXTBOOK	1.047	.250	1.071	.437
TWORKBK	-.434	.157	-.417	.275
TORDER1	-.023	.006	-.025	.010
Variance	38.271	6.186	-	-
Pupil-level Variance	-		36.138	-
Pupil-level Sigma	-		6.012	-
Group-level Variance	-		2.353	-
Group-level Sigma	-		1.534	.169
Deviance	-		19537.962	-

Table 4: Summary of Displays 1-5

OLS Variance	38.03	38.11	37.95	38.07	38.27
St. error	6.17	6.17	6.16	6.17	6.19
VCS Pupil-level Variance	36.81	36.96	35.87	35.87	36.14
Sigma	6.07	6.08	5.99	5.99	6.01
VCS Group-level Variance					
For G. mean	1.32	1.35	2.29	2.43	2.35
Sigma	1.15	1.16	1.51	1.56	1.53
St error for Sigma	0.19	0.19	0.17	0.17	0.17
Sample size					
Pupils	2076	2076	2804	2804	3025
Groups	60	60	80	80	86

Modelling of group-level variation (random slopes and random differences)

Simultaneously with reducing the fixed (regression) part of the variance component model for our data, we also need to explore extensions of the random part in order to obtain a better description of the group-level variation than the one offered by the group-level variance. We have concentrated first on reduction of the fixed part to a shorter list of explanatory variables because: (a) the school-level variation is rather small, and (b) in the models with complex description of variation, the fixed effect estimates and their standard errors differ very little from the obtained so far (Display 5).

In the variance component models fitted so far (Displays 1-5) the within-group regressions are assumed to be constant across groups, with exception of the intercept (position) which has a fitted variance of 2.35. More generally, the regression coefficients with respect to any of the pupil-level variables may be allowed to vary across the groups. These variables, selected from the variables included in the fixed part, form the random part of the model. The group-level variables are not considered for the random part, because within-group regressions with respect to such variables cannot be identified.

Variance component models closely resemble the models for analysis of covariance. The simple variance component models correspond to ANCOVA models with no interactions of covariates with the grouping factor. The (complex) variance component models with variable within-group regressions (slopes and/or differences) correspond to ANCOVA models with group x covariate interactions. The difference between the variance component and ANCOVA models is in the emphasis on description of variation as opposed to differences among the groups and in the assumptions of normality of the group effects in the former. The model specification in both models is analogous:

a, list of covariates (fixed part),

b, sublist of covariates which have interactions with the grouping factor (random part).

We now turn to modelling of the random part. For a continuous variable included in the random part the within-group regression slopes with respect to this variable are assumed to be randomly varying (and normally distributed) with an unknown variance. For a categorical variable included in the random part the within-group (adjusted) differences among the categories are normally distributed. We can consider the 'stereotype' group, for which the regression is given by the fixed part model (the average regression), and the regressions for the groups vary around this average regression. The deviations of the regression coefficients form a random sample (i.i.d) from a multivariate normal distribution. The components of the vector of deviations (for a group) cannot be assumed to be independent, and so their covariance structure has to be considered, but the variances of these deviations (or random effects) are of main interest.

Data with only a moderate number of groups, as is the case in this analysis, contain only limited information about variation, comparable to the limited information about interactions in models of analysis of covariance. Information about the covariance structure is usually even scarcer. Therefore, if a large number of variances are included in the random part (and estimated as free parameters) we can expect high correlations among the estimates—large estimated variances with large standard errors. Also, the number of covariances to be estimated grows rapidly with the number of variances, and many of the estimated correlations corresponding to these covariances are then close to +1 or -1. The variance matrix with these variances and covariances is not of full rank, and the random effects are linearly dependent. Therefore it is important to adhere to the principle of parsomony and seek the simplest adequate description for group-level variation. In selection of covariances to be estimated we use the guidelines set by Goldstein (1987) and Longford (1987).

Although model selection for the random part involves only pupil-level variables (inclusion/exclusion), it is more complex than the selection for the fixed part because constraints can be imposed also on the covariances. The most general variance component model would involve 17 variances (the number of regression parameters in Display 5) and 17x16/2=136 covariances. Fitting such a model is clearly not a realistic proposition, and so model selection has to proceed by building up the random part from simpler to more complex models.

In model selection for the random part we have proceeded in the following stages. For all the models we used the same fixed part as in Display 5. The estimates and standard errors for the regression parameters differed very slightly from those in Display 5 for all these models. The estimates and standard errors for the regression parameters differed very slightly from those in Display 5 for all these models. This fact justifies post poc our approach of first settling the fixed part and then proceeding with modelling of the random parts. First we fitted models with one pupil-level variable in the random part. Using the likelihood ratio test to compare the fitted model to the model with simple random part (Display 5) we selected the following variables: XROT, XAGE, YDESIRE AND YMOREED

The first three variables are ordinal, and associated with one variance each. The likelihood ratio (difference of deviances) for each of the three corresponding models was larger than 3. This is a very conservative criterion, since we prefer to err on the side of inclusion. There are two parameters - a variance (slope-variance) and a covariance (slope-by-intercept covariance) involved, but they are not free parameters since they have to satisfy the condition of positive definiteness. The distribution of the difference of the deviances is c_2^2 if the correlation corresponding to the covariance is smaller than 1 in modulus. The problem of negative variances is resolved by estimating the square roots of the variances (sigmas).

Next we fitted the VC model with these four variables in the random part, and simplified the random part by excluding variables and setting certain covariances to 0. The variance associated with the variable XAGE was very small (.00095) and its square root had a low t-ratio (.75), and so it could be constrained by 0 (excluded). That implies a constraint on all the covariances involving XAGE which are also set to 0. The three remaining variables and the intercept are represented by a 6x6 variance matrix; 6 variances and 15 covariances, almost as many parameters as in the fixed part. The fitted variance matrix is

Intercept		2.581					
XROT		.0143	.00558				
YMOREED	cat	.191	.0388	.812			
	cat 3	.519	.0439	.0621	1.032		
	cat 4	.384	.0354	-.0241	.261	1.032	
YDESIRE		.0863	-.0127	-.307	-.303	-.346	.667

The decrement in deviance, compared with the VCS model (Display 5) is only 13, hardly warranting addition of these 21 parameters in the model.

The software used provides standard errors for the square roots of the variances (sigmas, diagonal elements of the matrix) and for the covariances. The sigmas and their standard errors are:

			YMOREED			
	Intercept	XROT	cat2	cat3	cat4	YDESIRE
Sigma	1.607	.0747	.901	1.175	1.016	.828
St. Error	.176	.0261	.429	.451	.640	.295

The standard error for the covariances involving XROT and categories of YMOREED (rows 3-5 in column 2) are between .059 - .063 and for those involving YDESIRE and YMOREED (columns 3-5 in row 6) are .56 - .62. Each of these covariances have a small t-ratio, and so they were constrained to 0 in the next model. The following estimated variance matrix was obtained (the sigmas and their standard errors are given to the right of the variance matrix):

						Sigma	St. Error
2.415						1.554	.162
.0455	.00390					.0625	.0313
0	0	0				0	0
1.136	0	0	1.788			1.337	.341
.740	0	0	1.157	1.424		1.193	.514
.304	-.0436	0	0	.0	.830	.911	.260

The rank of this matrix is 4 (the two variance matrices given above are also singular), and so it would appear that another variance parameter could be constrained to 0. However, the t-ratio for each of the sigmas is high, and only a complex linear reparametrization of the variables included in the random part would enable further model simplification. The variance matrix obtained provides a description of group-level variation in terms of 11 parameters, 5 variances and 6 covariances. But the difference of variance of this model and the corresponding VCS model is only 11 (for 10 parameters). That provides further evidence of overparametrization or collinearity in the random part. However, any atempt to define a suitable model with fewer parameters would necessarily involve some unnatrually defined variables, which would deem interpretation of the model very difficult.

Variation in the slope of XROT provides evidence of unequal 'converison' of ability at the beginning of the year into ability at the end of the year. Such a conclusion is appropriate only subject to the caveats discussed in the Summary. The slope of XROT is shallower in some schools, where the initial differences in XROT tend to be associated with smaller differences in YROT than in schools where the slopes are steeper.

The regression slope for YDESIRE is about .5—this is the regression slope for the 'stereotype' school, where every feature is 'average'. The variation associated with this regression slope has a standard deviation of .9, and so there is a large (predicted) proportion of schools where the slope on YDESIRE is very small, or even negative! The correlation of the within-group slopes on XROT and YDESIRE is -.77; lower 'effects' of motivation to succeed are associated with schools where the initial differences become exaggerated by the end of the year.

The variances associated with the categories 3 and 4 of YMOREED represent the variation of the adjusted differences between categories 3 and 1 and 4 and 1, respectively. While the fitted difference between categories 2 and 1 is about .8, and constant for all the schools, the average within-school difference between categories 3 and 1 is 1.1, with a variance of 1.8. Therefore this difference is negative in several schools. The situation with the 4-1 contrast is similar, although the number of schools with reversed sign of the difference is much smaller. The correlation of the random effects associated with the categories 3 and 4 is .725 - high 3-1 contrast is associated with a high 4-1 contrast, but the fitted variance for the contrast 4-3 is 1.7? + 1.42 - 2*1.16 = .89, whereas the average difference is 1.58 - 1.08 - .50. Hence there are schools where the pupils with YMOREED = 3 have lower adjusted scores on YROT than YMOREED = 4, although on average the 4th category is .5 points ahead.

The estimates of the regression parameters differ only marginally for the different specifications of the random part. This justifies, post hoc, our approach of modelling first the regression part of the model and then the random part. The regression estimates for the last model considered are given in Display 6.

Display 6: Fixed-effect Estimates for the Final Model with Random Effects, for 3025 Students and 86 Classrooms/Schools Using 18 Explanatory Variagles Thailand 1981-82

Variable	VCS Estimate	VCS St. Error
GRAND MEAN	16.642	-
XROT	.617	.020
XAGE	-.070	.014
SXES	1.143	.260
YFOCCI	.101	.352
	-.488	.374
	.198	.446
YMEDUC	.347	.268
	.062	.446
	-.491	.560
YMOREED	.816	.453
	1.117	.476
	1.618	.514
YPARENC	.358	.112
YPERCEV	-1.178	.133
YFUTURE	.526	.137
YDESIRE	.480	.217
SENROLT	.300	.265
SPUTEAR	-.063	.048
SQUALMT	.781	.380
TMTHSUB	2.632	.582
TXTBOOK	0.949	.431
TWORKBK	-.372	.270
TORDER1	-.035	.270
TSEAT1	.007	.006
Variance	-	-
Pupil-level Variance	35.259	-
pupil-level Sigma	5.938	-
Group-level Variance	See matrix given in the text.	
Group-level Sigma		-
Deviance	19064.902	
Number of iterations	8	

<u>Conditional expectations of the random effects.</u>

In the fixed-effects ANOVA or ANCOVA, estimates of the effects associated with the groups are obtained. In variance component models these effects are represented by random variables. Conditional upon the adopted model the expectations of the (random) group-effects can be considered as the group-level residuals, or as "estimates" of the group-effects. These conditional expectations have to be inspected whether they conform with the assumptions of normality. This inspection involves a check for skewness and kurtosis (not carried out here, but visual inspection indicates no problems), and a chec' outlying values of the effects. The latter check is obviously also of substantive importance because it would be useful to detect schools with exceptionally high or low performance, where the categories of YMOREED have substantially different differences than the average school, in which schools the outcomes are more/less influenced by the initial score XROT. The complex nature of variation, involving three variables, coupled with the number of groups, makes it infeasible to discuss the deviations of the group-level regressions from the average regression. In fact, the main motivation in use of variance component analysis has been to obtrain a <u>global</u> description of variation, without reference to the individual groups. The added advantage is that owing to the shrinkage property of the conditional expectations extreme results due to unreliability for some of the schools with small numbers of students are avoided. The conditional expectations are a mixture of the pooled ordinary least squares solution of the within-group regression; the weight depends on the amount of information contained in the data from the group. Conditional expectations are obtained number of regression parameters. Owing to this shrinkage we cannot pinpoint to all the schools where (say) the difference of the categories 3 and 1 has a negative sign. For several schools the conditional means indicate a small difference between the categories; some of these may be negative, others positive and larger than the conditional expectation. Accordingly we should downscale our notion of what is an exceptionally large deviation; say, 1.5 multiple of the standard deviation (sigma) should be regarded as exceptional.

We conclude with an example of an exceptional school. School 22 (42 pupils in the data) has all its random-effects components positive. Its deviation from the average regression formula is

$$1.517 + .100\,\text{XROT} + .102\,\text{YDESIRE} + 1.008\,\text{YM}_3 + .842\,\text{YM}_4,$$

where YM_3 (and YM_4) are equal to 1 if the pupil is in category 3 (4), and 0 otherwise. This indicates that it is a school with high performance where the differences in initial ability tend to get exaggerated, pupils with high motivation and high expectations are at an advantage. For sample mean values of XROT and YDESIRE this formula becomes

$$2.959 + 1.008\, YM_3 + .842\, YM_4,$$

which reflects the high 'performance' of the school much more clearly. The variances quoted above refer to a regression using centred versions of all the variables

$$(XROT - \overline{XROT},\ YDESIRE - \overline{YDESIRE},\ YM_3 - \overline{YM_3},\ YM_4 - \overline{YM_4}).$$

In the transformation from one parametrization to the other only in the intercept-variance is affected.

DISCUSSION

At the outset of this paper, we posed three substantive and one methodological questions: (a) What characteristics of teachers and schools enhance student achievement?, (b) Are these effects uniform across different students?, (c) What is the comparative effectiveness of alternative inputs?, and (d) How do estimates obtained from simple OLS methods compare with estimates obtained from multilevel methods? During the development of the analysis, a fifth question arose: Are there alternative regression models that predict student achievement equally well as the model developed herein? In this section, we review our findings and present some caveats about their interpretations.

Summary

Effective teacher and school characteristics. The results from our final analysis indicate that there are teacher and school characteristics that are positively associated with student learning. These are:

- the percentage of teachers in the school that are qualified to teach mathematics,
- an enriched mathematics curriculum, and
- the frequent use of textbooks by teachers.

At the same time, some teaching practices are negatively related with learning; for this sample they are:

- the frequent use of workbooks, and

- time spent o maintaining order in the classroom.

The positive results are not surprising. Teachers who know the subject matter being taught, a curriculum that covers the domain, and textbooks that provide a structured presentation of the material all should have positive effects on achievement. The negative results are more curious. On the one hand, teachers who spend a great deal of time maintaining classroom order will have less time available for teaching; therefore, less learning takes place. On the other hand, the use of workbooks ought to contribute positively to achievement, not detract from it. Possibly the use of workbooks substitutes for something else: direct instruction, perhaps.

Uniformity of effects. In this sample, we found that schools did not have uniform effects on all students. In particular, effects differed according to the level of education expectations held by the students. Some schools/classrooms were more effective for students with low expectations, some were more effective for students with high expectations, while other schools are equally effective (or ineffective) for all types of students. Interestingly enough, we found little evidence that schools were differentially effective for students on the basis of sex, age, parental occupation or several other student attitudes. Thus, Thai schools were operating, by and large, in an egalitarian fashion, with the one exception of differences according to educational expectations.

Comparative effectiveness of inputs. Overall, we found few school "inputs" that were associated with differential achievement over time. Frequent use of textbooks increased achievement by a full point on the posttest, while use of workbooks decreased achievement by a third of a point; an enriched curriculum increased posttest scores by over 2.5 points. Each additional percentage of teachers that were qualified to teach mathematics raised posttest scores by over one point.

However, these causal statements do not hold if they are to be interpreted as a result of an external intervention. Obtaining (additional) textbooks for the schools is not a simple procedure unrelated to educational processes and management decisions; it is itself an outcome variable related to some (unknown) aspects of the educational process. Similarly, discarding workbooks would not lead to improved outcomes, unless all the circumstances that lead to reduced use of workbooks are also present, or are induced externally. External intervention will be free of risk only if we have, and apply, causal models for how the educational system functions. The models developed in this paper, and elsewhere in educational research literature, are purely descriptive. Use of regression methods, and of variance component analysis, allows improved description, but does not provide inference about causal relationships.

Also, interpretations of estimates of effects are subject to a variety of influences, and there may be alternative regression models, with different variables, that are equally correct in terms of prediction. Thus, the selection of variables included in this model is responsible, to some degree, for the results, and a different seleciton of variables could yield substantially different results with respect to the contribution of each variable.

Comparison with OLS. The analysis carried out demonstrates that estimates based on OLS regressions do yield different results, in some cases, to those based on VC regressions. For example, in comparing the OLS estimates with the VCS estimates in Display 6, we see that for TMTHSUB, the coefficients are quite different. Using OLS, we would conclude that students in "enriched" classes, controlling for the other explanatory variables, perform about 2 points (13%) higher than those in "normal" or "remedial" classes; the conclusion based on the VC regression is that they perform nearly 2.6 points (17%) higher. Combining these effect with cost information permits an estimation of cost-effectiveness. If enriched classes cost 13% more than remedial classes, we would conclude that they were either equally cost-effective (OLS) or more cost effective (VC) than remedial/normal classes, depending on the model. Similarly, if enriched classes cost 17% more than remedial/normal classes, they would be either equally cost-effective (VC) or less cost-effective (OLS), depending on the model. However, the caution about the causal inference in the previous subsection equally applies in this context. Classes, or schools, cannot be declared to have enriched curriculum at an external will and by supplying the outward signs of having enriched curriculum; rather, a whole complex of related circumstances have to be arranged, e.g., strengthened education in lower grades, synchronization with other subjects, etc. Since we have argued earlier in the paper that estimates based on VC methods are preferable to those based on OLS methods, differences of these types could hold important policy implications for schools deciding on the type of curriculum to choose.

Caveats

We have noted that alternative models could yield similar predictions (in terms of achievement), but might include a different set of variables. That such could be the case is not a problem limited to VC models; it is a perennial problem with these general types of analyses. In our analysis, we have included a number of individual pupil and school/classroom variables; in this respect, we have moved well beyond earlier models, which include only modest "intake" characteristics of students. Having identified the variables associated with higher outcome scores does not offer a direct answer to the principal question of a development agency about distribution of its resources to a set, or a continuum, of intervention policies in an educational system. Without any prior knowledge of the educational system, any justififcation for an intervention policy based on the results of regression (or variance component) analysis, or even of structural modelling (LISREL), would have no proper foundation. Certain intervention policies may cause a change in the educational system, and hence a change in the regression model itself. This new regression model may indicate that the selected intervention is far from optimal, or may even be detrimental.

A case in point is the pretest score XROT. Its coefficient is positive and of substantial magnitude. A conceivable intervention policy would be to raise the XROT scores, for example, by coaching prior to pretest administration. Clearly such an intervention, if effective, could lead to a change in the regression formula. Alternatively, if coaching took place between the pretest and posttest administrations, the regression formula would again be changed, out differently. Any number of different scenarios are easy to construct, in which the coefficient on XROT would be close to 1, or substantially lower than .62 (obtained in our analysis).

Similarly, indiscriminant reduction of the time spend on maintaining order in the classroom, probably a less expensive intervention in monetary terms, is likely to be an unreasonable solution. Introduction of the enriched mathematics curriculum for all students is most likely not practicable, and even its extension for a few more classrooms may place excessive requirements on staff in the schools, thus lowering the quality of instruction in other subjects, and/or other grades.

In conclusion, positive or negative regression coefficients cannot be regarded as indicators of cause, effect, or influence. An intervention could be regarded as an experiment, and its outcome can be predicted from an observational study only under the unrealistic assumptions of the regression formula describing accurately the mechanics of a rigid educational process.

Three important items of information would assist in answering the question about allocation of resources:

1. Feasibility and cost of various interventions.
2. How will an intervention effect other explanatory variables and which aspects of the educational process will remain unaltered after the intervention.
3. How directly manipulable are the "interventions"?

It is key to make distinction between variables that are manifest (unchangeable, e.g., pupil background), that are manipulable (e.g., time spent on a task of a particular kind), and that are manipulable only by direct intervention. For example, the time spend on maintaining discipline is a manipulable variable, but it can be either manipulated indirectly (e.g., by making the curriculum more interesting by providing more suitable or more interesting textbooks, or directly (through changing teacher behavior, so as to ignore disruptive student behavior). Effective education policy considerations require attention to directly manipulable variables; in the present analysis, these are the qualifications of the mathematics teachers in the school and the use of textbooks.

---

[1]These hierarchical structures result from design elements (stratified sampling), data collection technicalities (e.g., interviewer effect) or intrinsic interest in cross-level effects (e.g., the effects of post-natal feeding programs on the relationship between birth weight and subsequent cognitive development).

[2]An extended discussion of this is provided by H. Goldstein (1987).

[3]For more detail on the construction of the achievement measures, see Lockheed, Vail & Fuller, 1986.

References

Aitkin, M. & Longford, N. (1986). Statistical modelling issues in school effectiveness studies (with discussion). Journal of the Royal Statistical Society, Series B, 149, 1-43.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., & Congdon, Jr., R. T. (1986). An Introduction to HLM: Computer Program and User's Guide. Chicago: University of Chicago (processed).

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B, 39, 1-38.

Fuller, B. (1987). Raising school quality in developing countries: What investments boost learning? Review of Educational Research, 57, 255-291.

Goldstein, H. (1984). The methodology of school comparisons. Oxford Review of Education, 10, 69-74.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73, 433-56.

Goldstein, H. (1987). Multilevel Models in Educational and Social Research. New York: Oxford University Press.

Heyneman, S. (1986). The search for school effects in developing countries: 1966-1986. EDI Seminar Paper No. 33. Washgton, D.C.: World Bank.

Heyneman, S. P., & Amison, D. T. (1980). Student learning in Uganda: Textbook availablilty and other factors. Comparative Education Review, 23, 206-220.

Heyneman, S. P., Loxley, W. (1983). The effect of primary school quality on academic achievement across twenty-nine high and low-income countries. American Journal of Sociology, 88, 1162-1194.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society, Series B, 43, 1-41.

Lockheed, M. E., Fuller, B., & Nyirongo, R. (1987).. Family effects on student achievement in Thailand and Malawi. World Bank: Population and Human Resources Department (processed).

Lockheed, M. E., & Hanushek, E. (1988). Improving educational efficiency in developing countries: What do we know? Compare, 18, 21-38.

Lockheed, M. E., Vail, S, & Fuller, B. (1987). How textbooks affect achievement in developing countries: Evidence from Thailand. Educational Evaluation and Policy Analysis, 8, 379-392.

Longford, N. T. (1986). VARCL—Interactive software for variance component analysis. The Professional Statistician, 5, 28-32.

Longford, N. T. (9187). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika, 74, 817-827.

Mason, W. M., Wong, G. Y., & Entwistle, B. (1984). The multilevel linear model: A better way to do contextual analysis. Sociological Methodology. London: Jossey-Bass.

Psacharopoulos, G., & Loxley, W. (1986). Diversified Secondary Education and Development. Baltimore, MD: Johns Hopkins University Press.

Raudenbush, S. W. (1987). Educational applications of hierarchical linear models: A review. Michigan State University (processed).

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1-17.

Reynolds, D. (1985). Introduction: Ten years on—a decade of research and activity in school effectiveness research reviewed. In D. Reynolds (Ed.), Studying School Effectiveness. London: The Falmer Press.

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. Child Development, 54, 1-29.

Sirotnik, K. A., & Burstein, L. (1985). Measurement and statistical issues inmultilevel research on schooling. Educational Administration Quarterly, 21, 1699-185.

Willms, J. D. (1987. Differences between Scottish education authorities in their examination attainment. Oxford Review of Education, 13, 211-231.

INSTRUCTIONALLY SENSITIVE PSYCHOMETRICS: APPLICATIONS OF THE SECOND INTERNATIONAL MATHEMATICS STUDY

Bengt O. Muthén
CRESST and
Graduate School of Education
University of California, Los Angeles

1. Introduction

This paper discusses new psychometric analyses that improve capabilities for relating performance on achievement test items to instruction received by the examinees. The modeling discussion will be closely tied to data for U.S. eighth grade students provided by the Second International Mathematics Study (SIMS), comprising not only responses to a set of achievement items at the beginning and end of the eighth grade but also a relatively rich set of student background information, including opportunity-to-learn (OTL) information specific to each item (Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985).

Item Response Theory (IRT) is a standard psychometric approach for analyzing a set of dichotomously scored test items. Standard IRT modeling assumes that the items measure a unidimensional trait. This particular kind of latent trait model is used to assess the measurement qualities of each item and to give each examinee a latent trait score. As will be shown, however, IRT modeling is limited in ways that are a hindrance to properly relating achievement responses to instructional experiences. Taking IRT as a starting point, this paper summarizes the author's work on a set of new analytic techniques that give a richer description of achievement-instruction relations. Six topics that expand standard IRT and specifically deal with effects of varying instructional opportunities (OTL) will be discussed as outlined below.

1. Variation in latent trait measurement characteristics. This relates to the classic IRT concern of "item bias," here translated as the absence or presence of an added advantage due to OTL in getting an item right.

2. Multidimensional modeling. Inclusion of narrowly N defined, specific factors closely related to instructional units in the presence of a general, dominant trait.

3. Modeling with heterogeneity in levels. Analyses that take into account that achievement data often are not sampled from a single student population but one with heterogeneity of performance levels.

4. Estimation of trait scores. Deriving scores based on both performance and background information for both general and specific traits.

5. Predicting achievement. Latent trait modeling that relates to trait to student background variables.

6. Analyzing change. Relating change in general and specific traits to OTL.

The SIMS data will be used throughout to illustrate the new methods. All analyses will be carried out within the modeling framework of the LISCOMP computer program (Muthén, 1984, 1987).

Section 2 describes the SIMS data to be analyzed. Section 3 describes general features of the psychometric problem. Section 4 presents a descriptive analysis of the achievement N instruction relation for the SIMS data and sets the stage for later modeling. Sections 5-10 discuss methods topics 1-6 listed above.

2. The SIMS data

The Second International Mathematics Study (Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985) was conducted in order to study variations in mathematics knowledge for eighth and twelfth graders within and across several countries To this aim, multiple-choice mathematics achievement responses were collected on items in the areas of arithmetic, algebra, geometry, measurement, and statistics. The test was administered both in the Fall and in the Spring of each grade. The achievement test consisted of 180 items distributed among five test forms. Each student responded to a core test of 40 items and one of four randomly assigned rotated forms with about 35 items. For the part of the sample that we will be concerned with, the core test was administered both during the Fall and the Spring to all students in the study while the rotated forms varied in their use pattern. It is well known that eighth grade mathematics curricula vary widely, certainly for students in the U.S. To be able to better describe the variation in student math achievement, information related to these curricular differences was there .e also collected. A detailed part of this information was opportunity-to-learn (OTL) for the topics covered by each test item. For the U.S. eighth grade math students, information was also collected in order to make a distinction between "tracks" or class type, yielding a categorization into Remedial, Typical, Enriched, and Algebra classes. This classification was based on teacher questionnaire data and on information on textbooks used. A variety of other teacher-related information was also collected, such as topic emphasis, and teaching style. Student background information on family, career interests, and attitudes was also collected. We will concentrate our analysis on the U.S. eighth graders (for whom there are about 4,000 observations from both Fall and Spring) sampled from about 200 randomly sampled classrooms varying in size from about 5 to 35 students. We will be particularly concerned with analyses of the 40 core items, but will also report on analyses of the four rotated forms which, when combined with the core items, represent about 75 items administered to the about 1,000 students taking each form. The rotated form analyses will be presented as a cross-validation of findings for the core items. In this way, the SIMS data provide a uniquely rich set of data with which to study instructionally N sensitive psychometrics.

In the analyses that follow, a key piece of instructional information was obtained from the teacher questionnaire. For each item, teachers were asked two questions regarding opportunity to learn.

Question 1:

"During this school year did you teach or review the mathematics needed to answer the item correctly?"

1. No
2. Yes
3. No response

Question 2:

"If in the school year you did not teach or review the mathematics needed to answer this item correctly, was it mainly because?"

1. It had been taught prior to this school year
2. It will be taught later (this year or later)
3. It is not in the school curriculum at all
4. For other reasons
5. No response

Using these responses, opportunity-to-learn (OTL) level will be defined as;

No OTL: Question 1 (= 1), question 2 (= 2, 3, 4, or 5)

Prior OTL: Question (1 = 1, or 3) and question 2 (= 1)

This Year OTL: Question 1 (= 2), question 2 (= 9 (other response combinations had zero frequencies)

In most analyzes to follow, Prior OTL and this Year OTL will be combined into a single OTL category.

3. The General Problem

In general, psychometric modeling assumes independent and identically (i.i.d) distributed observations from some relevant population. This assumption is also made in IRT. The assumption of identically distributed observations is not realistic, however, using data of the SIMS kind to describe either relationships between what is measured (achievement responses) and what the measurements are attempting to capture (the traits), or how traits vary with relevant covariates such as instructional exposure and student background. This is because of the instructional heterogeneity of the students analyzed. The distribution of responses conditional on various traits values cannot be expected to be identical for a student who has had no specific instruction on the item topic and a student who has had instruction. The trait distribution cannot be expected to be the same for students in enriched classes as for students in typical classes. The students are naturally sampled from heterogeneous populations. It is true that increased homogeneity can be obtained by dividing the students into groups based on instructional experiences. However, such groupings may have to be very detailed to achieve their purpose and any simple grouping may be quite arbitrary. A more satisfactory approach is to use modeling that allows for heterogeneity, using parameters that vary for varying instructional experiences. Such modeling also accomplishes the goal of instructionally sensitive psychometrics, namely explicitly describing the achievement response-instructional experiences relations.

4. Descriptive analyses

Its informative to consider descriptively how the achievement responses vary with instructional exposure. This forms a basis for our subsequent modeling efforts. We will study this in terms of both univariate and bivariate achievement distributions using the posttest core items administered to the U.S. eighth graders. We will also study the change in univariate responses from pretest to posttest.

4.1 Univariate response

Consider first the univariate responses for the posttest. The wording of the core items is given in the appendix. The proportion correct for each item is described in Table 1, broken down by the class type categories Remedial, Typical, Enriched, and Algebra and by the OTL categories No OTL, This Year OTL, and Prior OTL. From the totals it is seen that both class type and OTL have a strong effect on proportion correct.

For most items the proportion correct is higher for Enriched and Algebra classes than for Remedial and Typical classes. For almost all items the proportion correct increases when moving from No OTL to this Year OTL to Prior OTL. The reason why Prior OTL gives higher proportion correct than This Year OTL is partly because Prior OTL is more common for Enriched and Algebra classes to which we presume students of higher achievement levels have been selected. OTL appears to have an overall positive effect on proportion correct also when controlling for class type, at least for typical classes. Also, when controlling for OTL, class type seems to still have a strong effect.

These univariate relationships are informative but confound effects of instructional exposure with effects of student achievement level. For example, the higher proportion

TABLE 1

Percentage Students and Percentage Correct for Core Items by OTL and Class Type

Item	Total*		No OTL			This Year OTL			Prior OTL		
	PR	PO	ST	PR	PO	ST	PR	PO	ST	PR	PO
ME01											
TOT	35	43	21	22	26	59	36	47	20	[illegible]	48
REM	11	18	33	7	8	60	12	23	7	21	21
TYP	30	38	42	21	27	64	34	43	12	28	34
ENR	42	52	17	25	24	71	48	63	12	29	29
ALG	61	64	6	64	64	5	39	50	89	62	65
AR02											
TOT	47	60	3	34	53	89	45	59	8		78
REM	12	21	9	17	33	91	11	20	0	0	0
TYP	42	57	3	34	40	97	42	57	6	74	81
ENR	58	74	4	46	86	90	57	73	6	74	81
ALG	74	75	0	0	0	43	73	71	57	74	78
AL03											
TOT	9	21	38	8	9	61	10	28	1	3	19
REM	15	9	78	15	8	22	13	13	0	0	0
TYP	8	14	49	7	9	50	8	18	2	3	19
ENR	8	21	16	12	11	84	7	23	0	0	0
ALG	16	64	7	0	19	94	17	68	0	0	0
AR04											
TOT	27	[illegible]	13	23	26	75	26	31	12	44	50
REM	16	33	40	16		60	16	15	0	0	0
TYP	24	29	11	16	[illegible]	87	25	30	2	30	30
ENR	29	38	15	39	48	70	25	34	15	37	45
ALG	47	54	0	0	0	33	41	50	67	50	56
ME05											
TOT	32	44	7	32	30	86	31	45	6	46	55
REM	17	18	6	27	27	85	17	18	9	17	8
TYP	27	40	8	20	17	90	27	42	2	22	43
ENR	37	55	5	49	60	95	37	54	0	0	0
ALG	56	63	8	75	66	48	53	62	44	55	64
ME06											
TOT	49	55	28	48	54	59	48	55	13	52	59
REM	20	31	41	23	35	45	21	31	14	11	22
TYP	47	52	27	48	53	65	48	52	8	42	47
ENR	52	61	32	51	60	65	52	62	2	82	68
ALG	66	73	10	83	80	28	68	75	62	63	72
GE07											
TOT	56	66	69	55	66	23	56	66	8	63	73
REM	26	39	75	25	36	25	27	46	0	0	0
TYP	54	64	66	54	64	24	55	63	10	56	67
ENR	58	72	71	56	71	27	64	75	3	62	77
ALG	77	85	75	76	84	6	82	95	19	83	87

TABLE 1
Percentage Students and Percentage Correct for Core Items by OTL and Class Type

Item	Total*		No OTL			This Year OTL			Prior OTL		
	PR	PO	ST	PR	PO	ST	PR	PO	ST	PR	PO
ME08											
TOT	89	89	17	89	88	58	88	88	25	93	92
REM	67	61	34	62	55	58	69	64	8	76	67
TYP	89	89	17	94	93	66	88	89	18	89	88
ENR	93	93	16	90	91	59	93	93	26	96	94
ALG	98	97	14	96	100	12	96	98	74	99	97
ME09											
TOT	42	52	14	41	48	56	38	50	30	50	59
REM	16	18	27	18	19	58	15	18	15	21	15
TYP	37	48	14	41	49	62	36	47	23	38	49
ENR	48	64	11	42	53	63	46	65	27	56	65
ALG	67	73	12	76	78	2	56	33	85	66	73
GE11											
TOT	26	31	40	20	26	56	29	34	4	33	39
REM	9	8	77	11	7	19	4	10	4	0	27
TYP	20	27	43	16	25	54	22	29	2	23	21
ENR	31	38	29	31	36	68	32	38	3	15	35
ALG	57	54	24	49	46	62	59	55	14	57	59
AR12											
TOT	34	44	10	32	40	85	34	44	5	41	48
REM	18	22	35	19	23	65	19	21	0	0	0
TYP	30	40	6	22	29	90	31	41	4	25	35
ENR	39	51	9	51	65	89	38	49	3	43	57
ALG	54	62	16	46	57	63	55	64	21	56	58
AL13											
TOT	58	71	12	46	59	85	59	73	2	74	85
REM	31	46	32	28	36	68	33	51	0	0	0
TYP	54	67	45	48	62	84	55	67	1	68	91
ENR	63	81	2	94	94	94	62	81	4	69	94
ALG	87	89	7	46	77	88	90	92	6	87	65
AR14											
TOT	56	61	15	49	53	78	56	61	7	66	76
REM	29	26	29	27	23	64	32	27	7	17	28
TYP	53	58	15	46	50	82	54	58	4	62	79
ENR	61	70	13	59	70	85	62	70	2	35	65
ALG	77	82	8	97	88	51	75	81	41	76	81
AR15											
TOT	22	32	10	20	28	77	20	30	14	34	45
REM	18	18	10	22	15	90	17	18	0	0	0
TYP	20	28	12	18	26	83	20	28	5	28	31
ENR	21	38	8	26	39	83	21	39	9	15	28
ALG	38	47	0	0	0	23	23	25	77	42	54
AL16											
TOT	23	58	6	9	16	92	24	60	2	37	88
REM	9	14	52	10	9	43	7	20	0	0	0
TYP	18	50	3	6	11	97	18	52	0	0	0
ENR	28	74	2	17	89	94	28	73	4	34	94
ALG	53	89	0	0	0	94	53	89	6	41	77

TABLE 1
Percentage Students and Percentage Correct for Core Items by OTL and Class Type

Item	Total*		No OTL			This Year OTL			Prior OTL		
	PR	PO	ST	PR	PO	ST	PR	PO	ST	PR	PO
GE17											
TOT	47	59	13	39	38	72	46	62	15	[illegible]	63
REM	24	24	41	22	15	48	25	26	10	29	46
TYP	42	56	11	42	37	82	43	60	8	35	40
ENR	53	68	12	44	44	80	55	72	[illegible]	53	68
ALG	76	80	10	61	85	18	78	93		78	77
AL18											
TOT	43	51	20	32	29	78	46	56	2	58	60
REM	25	23	55	20	17	45	31	31	0	0	0
TYP	39	44	24	36	31	76	40	48	0	0	0
ENR	47	63	4	28	36	89	47	65	6	59	57
ALG	71	78	7	31	58	98	75	81	6	55	68
GE19											
TOT	23	33	76	23	32	23	22	38	1	52	57
REM	10	19	0	10	19	0	0	0	0	0	0
ENR	25	39	71	25	35	29	25	49	0	0	0
ALG	39	49	89	38	48	0	0	0	11	52	57
AR20											
TOT	73	77	2	55	60	86	71	76	12	89	90
REM	31	37	7	28	33	93	31	37	0	0	0
TYP	71	75	3	64	69	93	71	75	5	78	85
ENR	80	87	6	0	88	94	80	87	0	84	0
ALG	94	94	0	0	0	29	93	96	71	94	93
GE21											
TOT	20	34	60	20	30	37	21	39	3	23	39
REM	16	16	97	16	17	3	25	13	0	0	0
TYP	18	30	60	17	29	39	20	33	1	22	11
ENR	20	39	46	20	34	52	20	44	2	6	11
ALG	34	50	65	33	45	18	44	71	17	28	49
GE22											
TOT	37	59	13	26	26	80	37	64	7	62	67
REM	21	18	79	23	19	17	9	11	4	30	40
TYP	33	55	8	28	26	90	33	58	2	29	37
ENR	40	71	6	20	15	92	40	75	2	59	59
ALG	70	81	9	47	82	44	70	85	47	73	78
ME23											
TOT	33	47	19	25	30	73	33	50	8	47	65
REM	17	18	52	17	18	48	18	17	0	0	0
TYP	29	41	19	25	30	80	31	44	2	16	29
ENR	33	58	15	29	41	79	34	62	6	23	53
ALG	59	74	7	38	35	43	62	78	51	60	76
AR24											
TOT	52	59	7	37	36	83	50	55	10	78	81
REM	23	18	15	33	18	85	21	18	0	0	0
TYP	47	53	10	38	40	89	48	55	1	50	38
ENR	60	66	0	0	0	95	60	65	5	61	75
ALG	80	82	0	0	0	21	71	76	79	82	83

TABLE 1

Percentage Students and Percentage Correct for C[illegible] Items by OTL and Class Type

Item	Total*		No OTL			This Year OTL			Prior OTL		
	PR	PO	ST	PR	PO	ST	PR	PO	ST	PR	PO
AL25											
TOT	42	46	7	28	34	92	42	47	2	70	59
REM	12	15	28	8	13	72	13	16	0	0	0
TYP	38	42	7	36	40	92	37	43	2	68	44
ENR	48	55	3	40	60	97	49	55	0	0	0
ALG	69	67	0	0	0	94	69	66	6	73	86
AL27											
TOT	46	57	53	38	50	47	54	64	1	67	71
REM	27	30	92	26	30	9	36	24	0	0	0
TYP	42	52	58	37	48	41	49	59	1	67	71
ENR	50	63	46	49	64	51	50	62	0	0	0
ALG	69	82	7	50	65	93	71	83	0	0	0
AR28											
TOT	51	62	9	44	49	74	49	61	16	63	73
REM	20	29	20	19	19	76	21	33	4	0	18
TYP	47	57	11	47	49	80	47	58	9	44	59
ENR	59	72	6	56	79	83	58	71	11	61	69
ALG	77	86	0	0	0	25	73	85	75	78	86
ME29											
TOT	77	75	10	63	60	64	76	75	25	83	81
REM	40	44	22	40	22	68	41	49	11	34	55
TYP	75	74	9	65	69	71	75	74	19	78	74
ENR	85	82	13	71	64	71	87	84	16	85	88
ALG	L92	89	0	0	0	11	95	95	89	91	89
AL30											
TOT	31	40	52	28	36	45	34	48	3	34	43
REM	25	23	83	27	23	17	13	20	0	0	0
TYP	27	37	59	23	35	38	28	40	3	39	29
ENR	34	46	37	32	41	57	35	48	6	28	58
ALG	50	57	25	48	61	75	51	56	0	0	0
AR33											
TOT	45	50	5	34	33	87	44	49	[illegible]	62	66
REM	20	19	22	20	12	78	20	21	0	0	0
TYP	41	47	5	39	41	91	41	47	4	52	57
ENR	51	59	0	0	0	97	50	59	3	74	61
ALG	65	69	2	75	75	47	65	67	51	65	71
AR34											
TOT	24	39	4	16	19	90	22	39	7	45	53
REM	10	15	19	14	16	81	9	14	0	0	0
TYP	19	34	4	17	22	96	19	34	0	0	0
ENR	29	54	0	0	0	97	29	54	3	39	35
ALG	44	53	0	0	0	43	53	50	57	45	55
AL35											
TOT	51	59	29	39	44	70	55	65	1	54	92
REM	38	30	78	37	33	22	41	22	0	0	0
TYP	46	55	36	40	46	63	49	59	1	54	92
ENR	53	68	11	37	52	89	55	70	0	0	0
ALG	78	83	0	0	0	100	78	83	0	0	0

TABLE 1
Percentage Students and Percentage Correct for Core Items by OTL and Class Type

Item	Total*		No OTL			This Year OTL			Prior OTL		
	PR	PO	ST	PR	PO	ST	PR	PO	ST	PR	PO
AR36											
TOT	47	56	7	44	38	86	46	56	7	64	73
REM	33	31	19	37	31	81	32	30	0	0	0
TYP	44	52	8	47	41	92	44	53	0	0	0
ENR	51	66	4	41	32	93	52	68	3	43	57
ALG	66	72	0	0	0	43	65	68	57	66	75
AR37											
TOT	31	37	15	21	23	65	29	36	21	44	52
REM	14	12	38	11	8	62	16	14	0	0	0
TYP	26	31	17	24	24	73	27	33	9	28	32
ENR	36	46	6	19	30	62	36	48	32	40	46
ALG	57	69	5	39	67	24	49	63	71	61	71
AR38											
TOT	36	51	3	26	23	91	34	51	7	61	72
REM	16	25	9	25	17	91	16	25	0	0	0
TYP	31	45	3	27	25	97	31	46	0	0	0
ENR	42	66	0	0	0	97	43	66	3	35	52
ALG	61	0	0	0	43	57	62	57	63	74	
GE40											
TOT	35	47	47	33	41	50	37	52	3	52	56
REM	24	31	93	24	31	7	21	21	0	0	0
TYP	32	43	46	30	38	54	34	46	0	0	0
ENR	39	56	32	35	44	66	42	63	2	22	50
ALG	52	60	56	53	59	19	44	68	26	57	57

* Percentage of students by class type are:
REM = Remedial: 7.1 (N=268), TYP=Typical: 57.6 (N=2148)
ENR=Enriched: 24.4 (N=909), ALG=Algebra: 10.7 (N=399)

ST=Percentage students
PR=Percentage correct for pretest
PO=Percentage correct for posttest

ME=measurement
AR=Arithmetic
AL=Algebra
GE=Geometry

correct for a certain item for students with Prior OTL may be solely due to such students having a higher achievement level on the whole test. It would be of interest to know if students with the same achievement level perform differently on a certain item for different instructional exposure. To this aim, we may consider the total score on the posttest as the general mathematics achievement level of each student and study the variation of proportion correct for each item as a function of instructional exposure conditionally on the general achievement level. We have carried this out using the dichotomous version of OTL, combining Prior OTL with This Year OTL into a single OTL category.

For each value of the achievement variable we then have a proportion correct for a No OTL and an OTL group and can study whether OTL makes a difference. Conversely, for each of the two OTL categories we will present the distribution of the achievement variable in order to study whether having OTL for an item implies that these students have a higher general achievement level. These plots are given in Figures 1-9.

Figure 1 describes items 1, 2, and 3. The left-most panel shows the total score distribution given No OTL and OTL, respectively. We note that the score distributions have different locations with the OTL distribution having somewhat higher mean, supporting the notion that students who receive OTL perform better as measured by this test. We also note that the variances of the two distributions are about the same. The score distributions shown are representative of all core items.

The right-most part of Figure 1 and Figures 2 - 9 contain curves showing the proportion correct for given total score for the two OTL categories. For each item and both OTL categories, proportion correct increases with total score indicating that for both OTL categories the item is a good indicator of the general achievement variable which the total score represents. It is particularly noteworthy that this is true also for the No OTL category and that the No OTL and OTL curves most often are very close. The students who, according to their teachers, have not been taught the mathematics needed to answer the item correctly still appear to have a high probability of answering the item correctly and this probability increases with increasing total score. This may indicate that students can to a large degree draw on related knowledge to solve the item. It may also indicate unreliability in the teachers' OTL responses. However, the differences in score distributions for the core items show that the OTL measures have consistent and strong relations to the total score. Instead of unreliability there may be a component of invalidity involved in the teachers' responses, where OTL may to some extent be confounded with average achievement level in the class and/or the item's difficulty.

The score distributions show that OTL is correlated with performance. Our hypotheses is that OTL helps to induce an increased level of general achievement variable and that in general it is this increased level that increases the probability of a correct answer, not OTL directly. In this way, moving from the No OTL status to the OTL status implies a move upwards to the right along the common curve for No OTL and OTL.

FIGURE 1

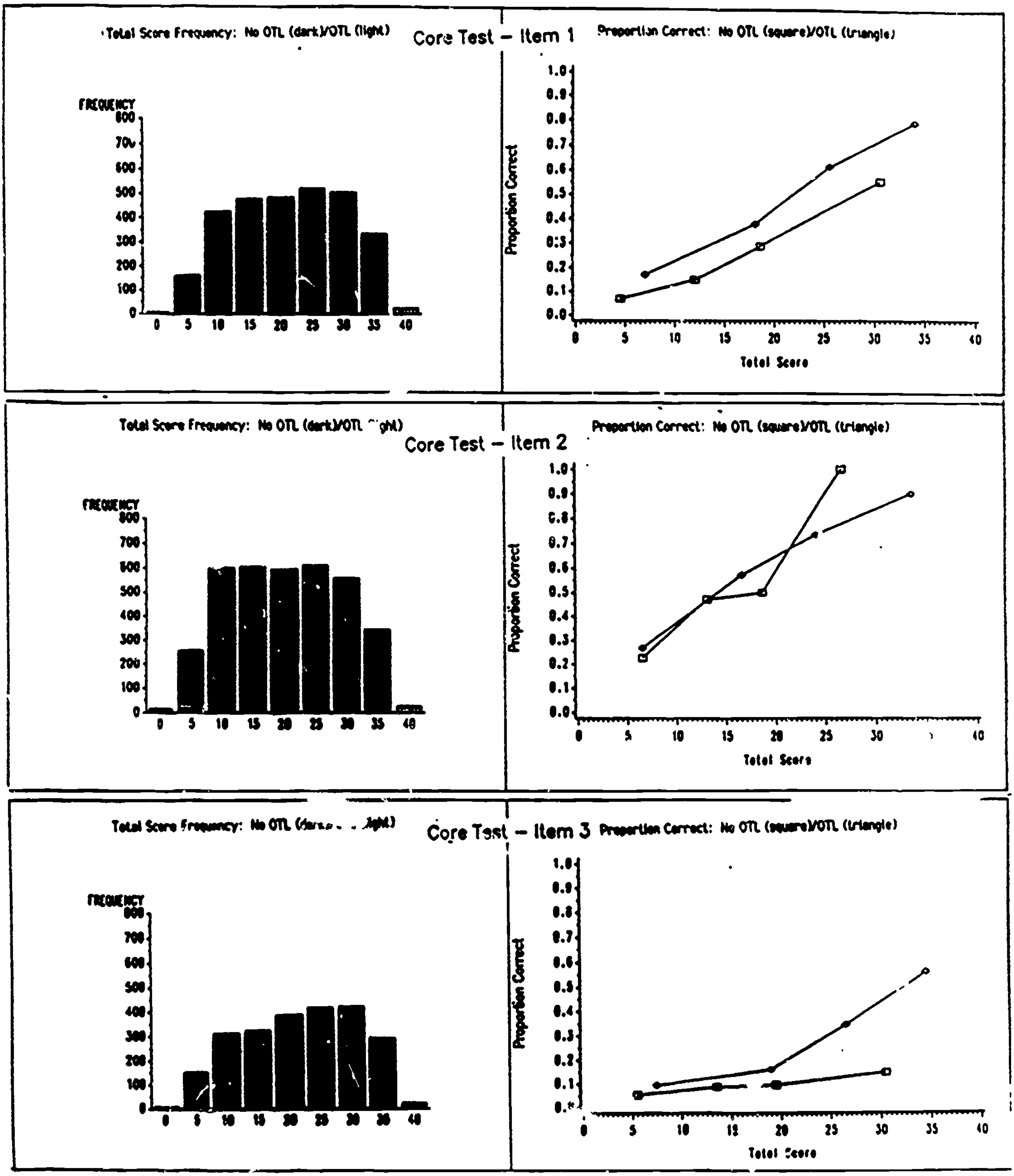

Total Score Frequency: No OTL (dark)/OTL (light)
Core Test – Item 1
Proportion Correct: No OTL (square)/OTL (triangle)
FREQUENCY
Proportion Correct
Total Score
Core Test – Item 2
Core Test – Item 3

FIGURE 2

Proportion Correct: No OTL (square)/OTL (triangle)

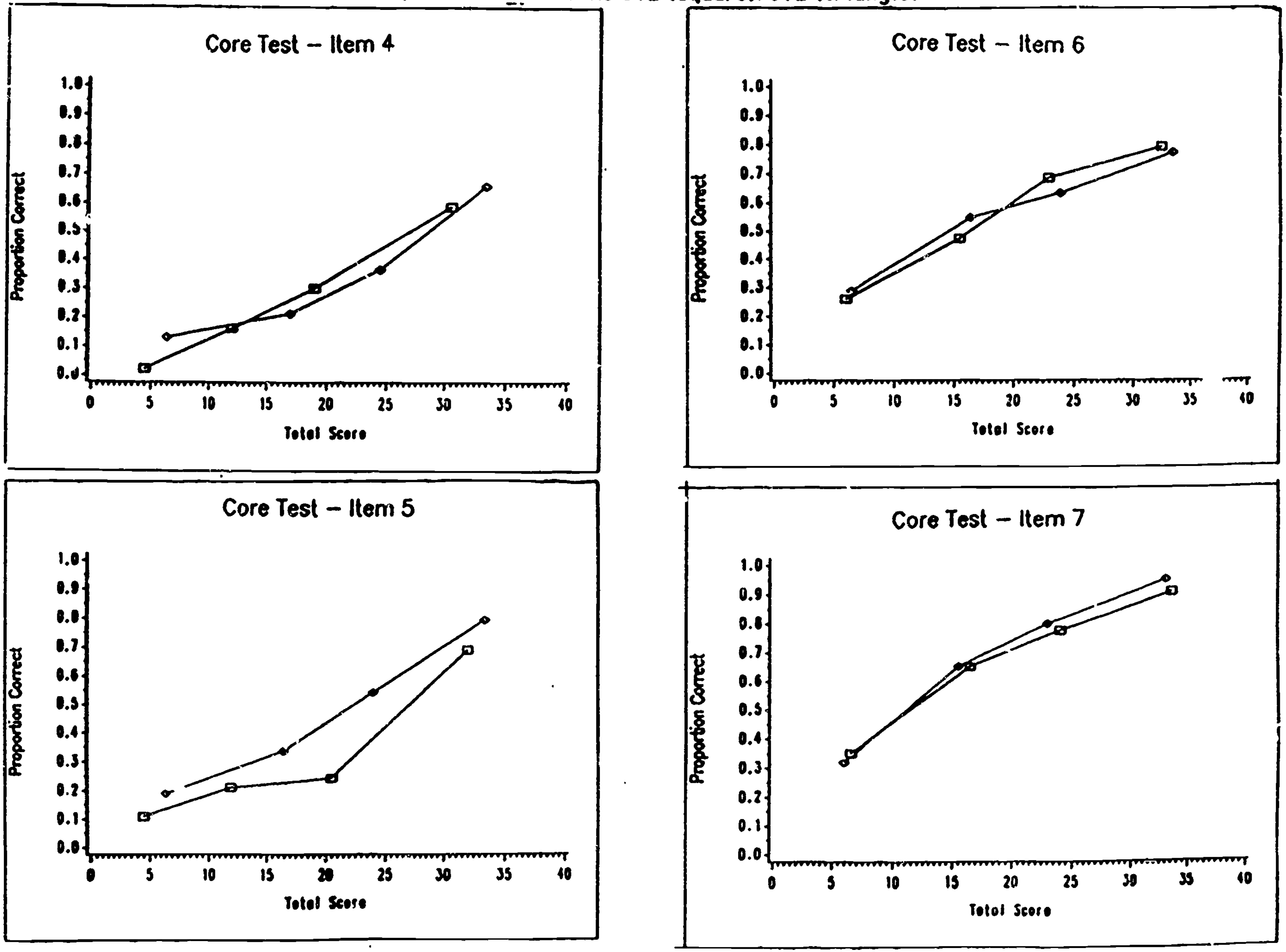


FIGURE 3

Proportion Correct: No OTL (square)/OTL (triangle)

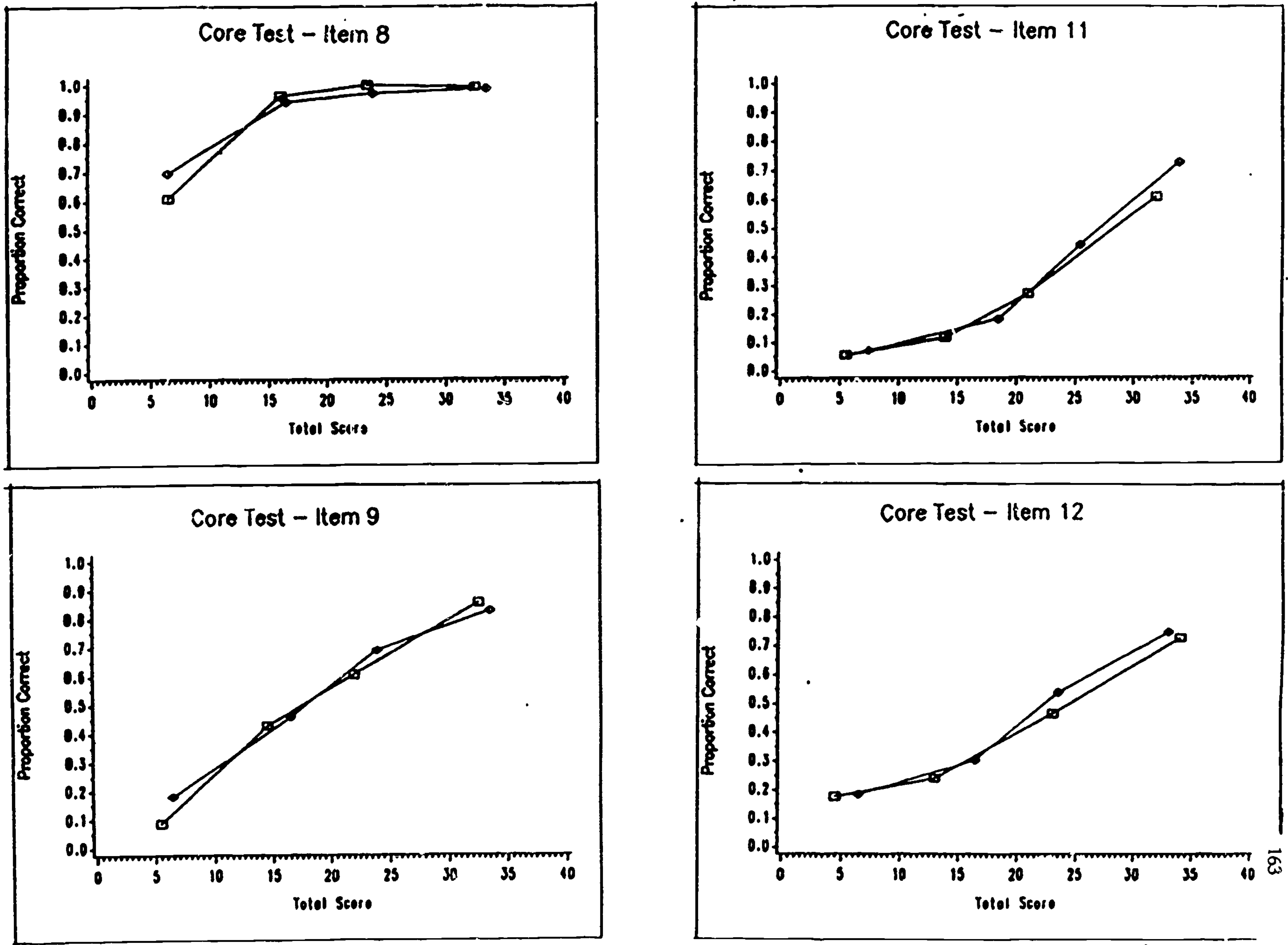


FIGURE 4

Proportion Correct: No OTL (square)/OTL (triangle)

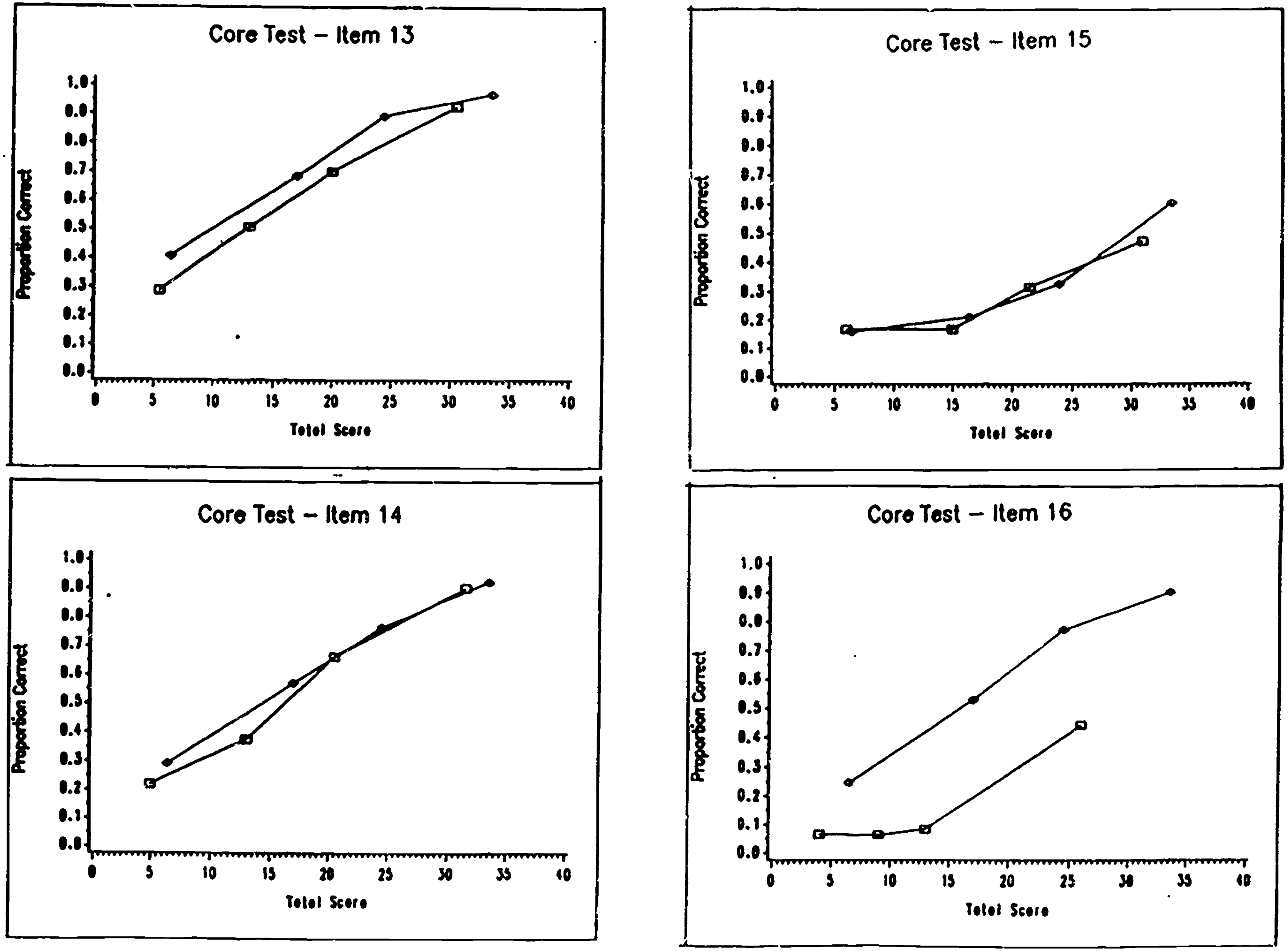


FIGURE 5

Proportion Correct: No OTL (square)/OTL (triangle)

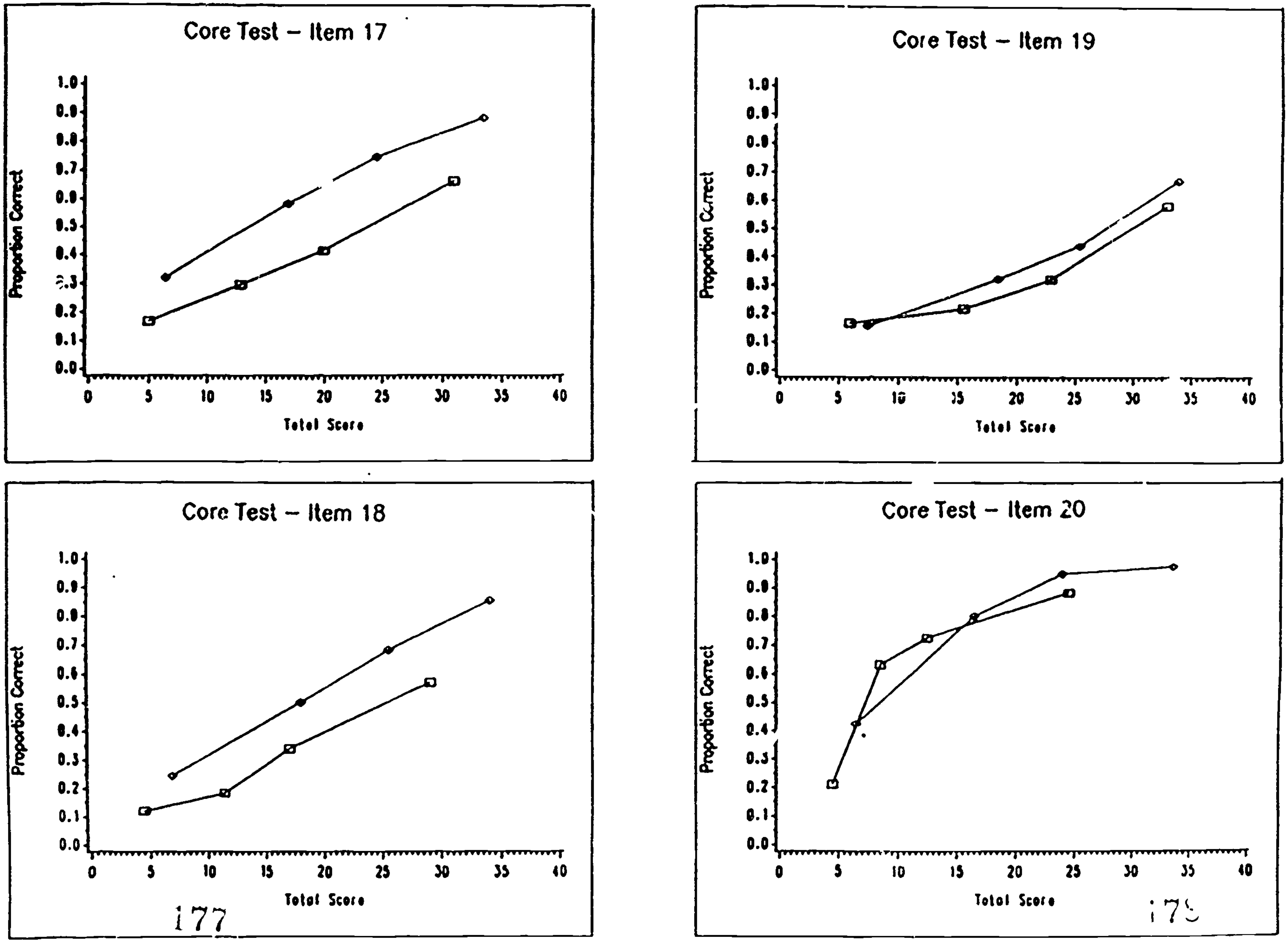


FIGURE 6

Proportion Correct: No OTL (square)/OTL (triangle)

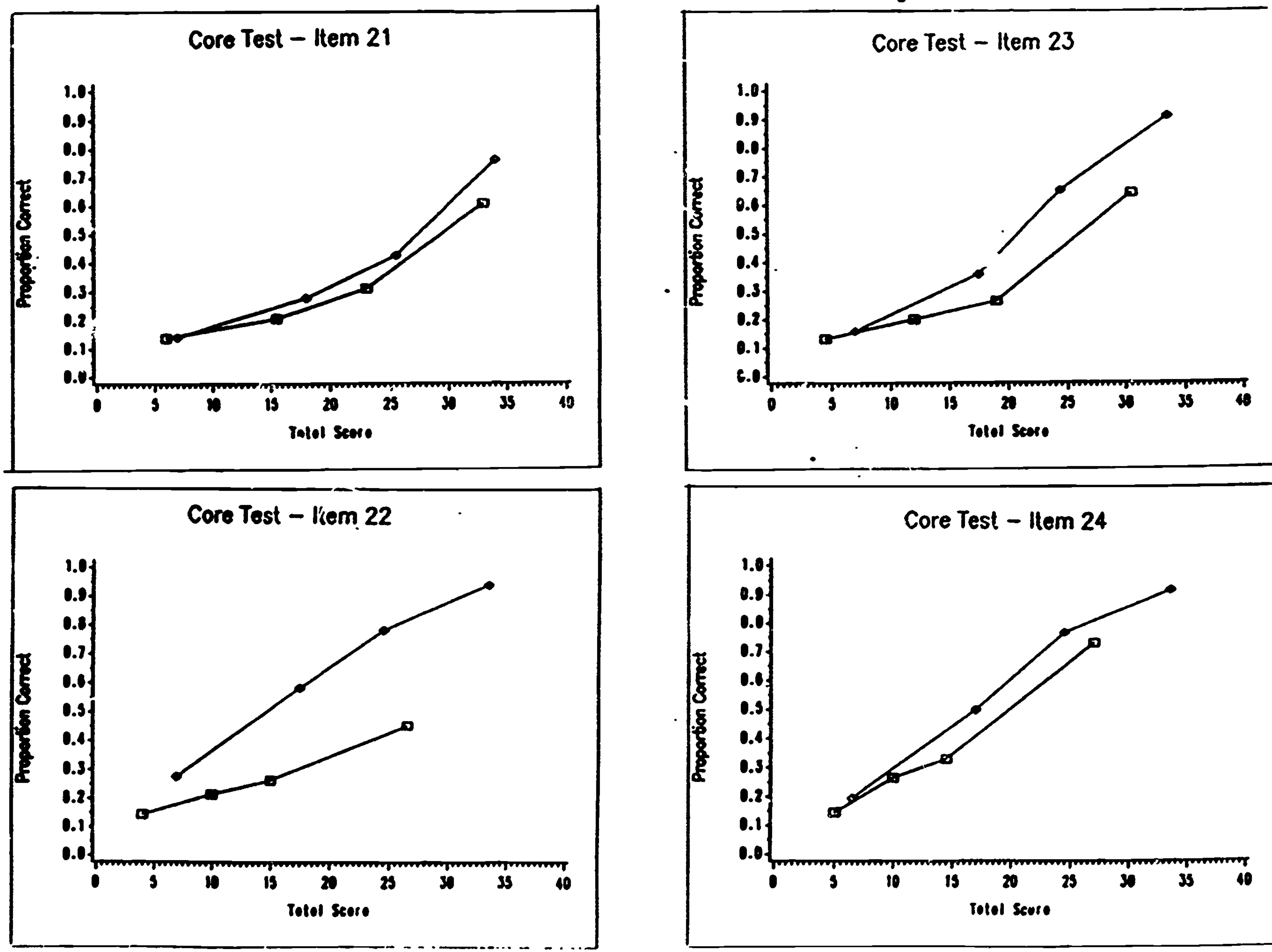


FIGURE 7

Proportion Correct: No OTL (square)/OTL (triangle)

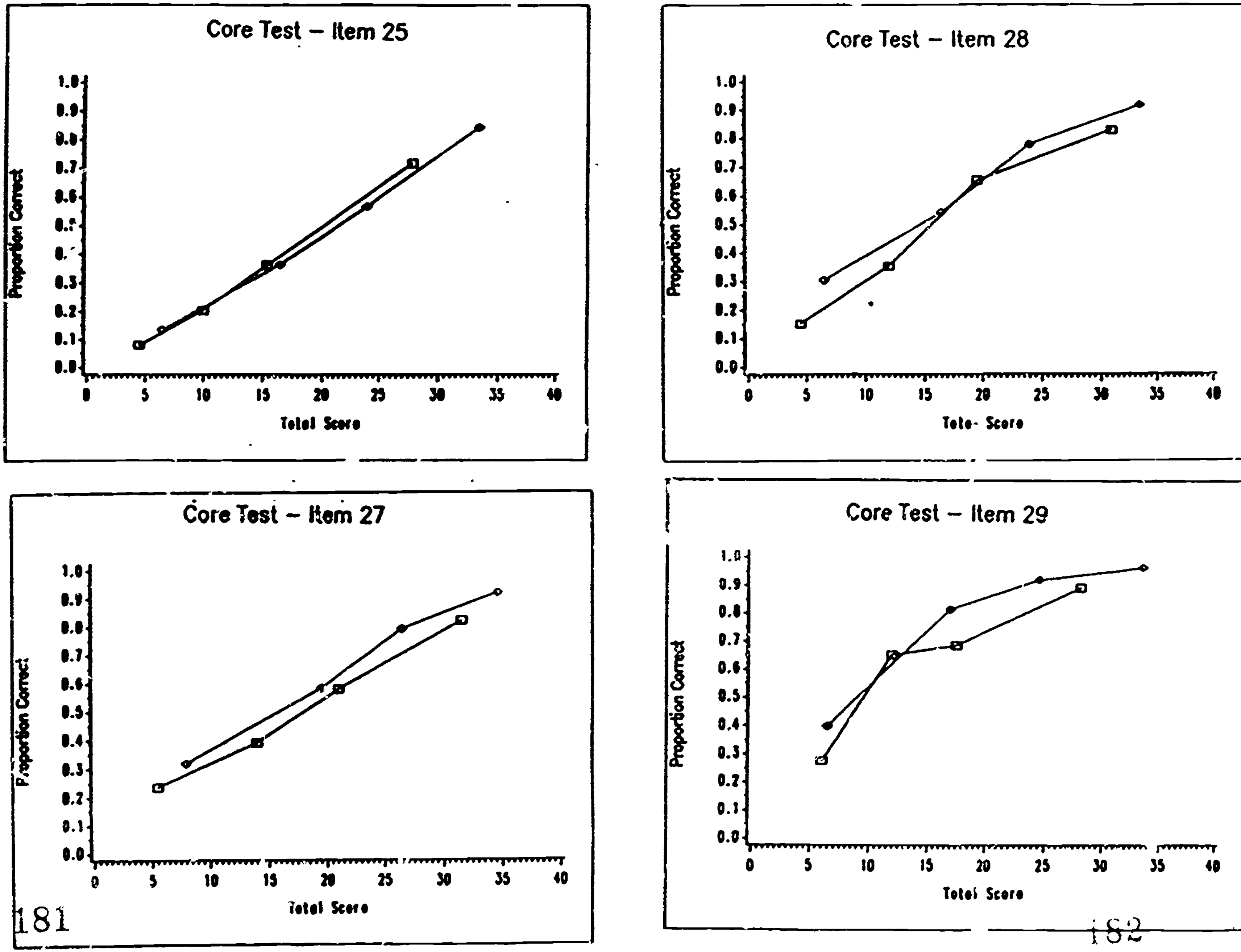


FIGURE 8

Proportion Correct: No OTL (square)/OTL (triangle)

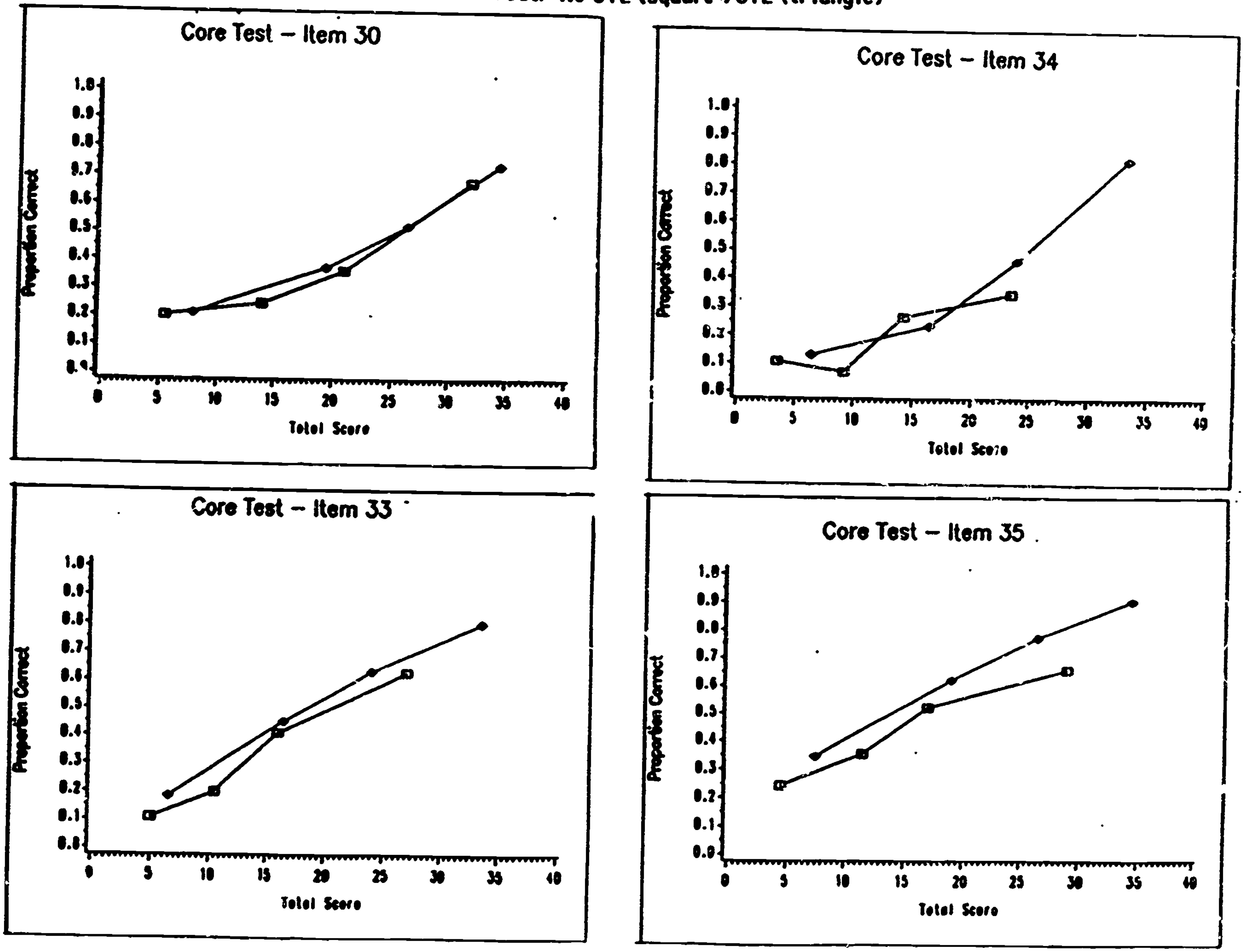
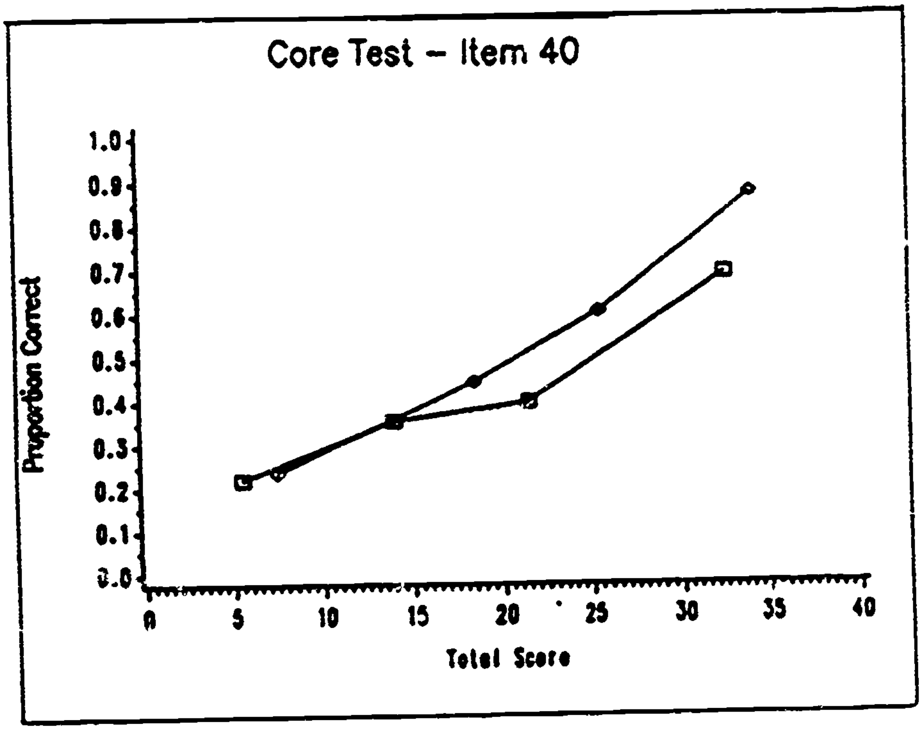


FIGURE 9

Proportion Correct: No OTL (square)/OTL (triangle)



There are some exceptions to the general finding of common curves for the No OTL and OTL categories. For example, items 3, 17, and 39 show a large positive effect of having OTL. Several other items with sizeable numbers of students in the two OTL categories also show positive effects. This means that for these items, the added advantage of having OTL is not fully explained by a corresponding increase in total score. OTL directly affects the success in solving the item correctly. From Table 1 we find that for the three items listed, the proportion correct increases strongly when moving from the No OTL category to the OTL categories. However, Table 1 cannot be counted on for finding items with direct OTL effects of this kind, since several other items also show strong increases in proportion correct due to OTL. We will return to the interpretation of this type of effect in Section 4. Note also that with the exception of item 3 an, OTL effect appears to be such that the two curves are approximately parallel, implying that the OTL effect is constant across achievement levels. For item 3 the OTL advantage increases with increasing achievement level, perhaps because it is a difficult item.

4.2 Bivariate responses

The various descriptive analyses carried out for the univariate responses can be carried over to bivariate responses. A common measure for studying relationships among dichotomous items is that of the tetrachoric correlation coefficient (Lord & Novick, 1968). In line with the previous section, we may study the strength of association between each pair of achievement items by computing three sets of correlations, using all students, students with No OTL on neither of the pair of variables, and students with OTL on both of the pair of variables. For each of the sets, the average correlation across all pairs gives an indication of the degree of homogeneity of the items in their measurement of achievement. It is of interest to study if this homogeneity is affected by OTL. Further, in line with the previous section, the corresponding three sets of correlations may be computed conditional on the total test score viewed as a general achievement variable. For lack of space these analyses will not be presented here, except to note that the homogeneity of correlatons does not seem to be affected by OTL.

4.3 Change of univariate responses

The SIMS core items also provide the opportunity to study changes in proportion correct for each item from the Fall testing to the Spring testing. This change can be related to OTL. For each item we may distinguish between three groups of students, those who did not have OTL before the pretest or before the posttest (the No OTL group), those who had OTL before the pretest (Prior OTL), and those who did not have OTL before the pretest but did have OTL before the posttest (This year OTL). The change for the No OTL group gives an indication of change due to learning on related topics. The change for the Prior OTL group gives an indication of effects related to practice, review, and, perhaps, forgetting. The change for the group having This Year OTL reflects the direct exposure to the topic represented by the item. These changes can be studied in Table 1. Table 1 shows that, where changes occur, they are largely positive for each OTL category with the largest changes occurring for students in the category of This Year OTL as expected. They may be taken to support the dependability of the teacher-reported OTL measure.

5. Variations in latent trait measurement characteristics

The study of the univariate achievement responses in Section 4.1 showed that the set of core test items served as good indicators of the total test score. We may hypothesize that this test score is a proxy for a general mathematics achievement variable as measured by the combined content of the set of core items. However, the total test score is a fallible measure and what we are interested in are the relationships between the items and the true score and estimates of the true scores. This is a situation for which Item Response Theory (IRT) has been proposed as a solution used (see for example, Lord, 1980). The curves of Figures 1 - 9 are, in IRT language, empirical item characteristic curves, which as theoretical counterparts have conditional probability curves describing the probability correct on an item given a latent trait score. We will now describe the IRT model and how it can be extended to take into account instructional heterogeneity in its measurement characteristic.

In formulas the IRT model may be briefly described as follows. Let y^* be a p vector of continuous latent response variables that correspond to specific skills needed to solve each item correctly for item j,

(i) $y_j = O$, if $y^* £ t_j$
i, otherwise

where O denotes the incorrect answer, 1 denotes the correct answer, and t_j is a threshold parameter for item j corresponding to its difficulty. Assume also that the latent response variable y^*_j is a function of a single continuous latent h and a residual e_j,

(2) $y^*_j - l_j h + e_j$.

where l_j is a slope parameter for j, interpretable as a factor loading. With proper assumptions on the right-hand-side variables, this gives rise to the two-parameter normal ogive IRT model. For each item there are two parameters t_j and l_j. The conditional probability of a correct response on item j is

$$(3)\ P(y_j = i \mid h) = F[(-t_j + l_j h)\, q^{-\frac{1}{2}}]$$

where q is the variance of e_j. This means that the threshold t_j determines the item's difficulty, that is the horizontal location of the probability curve, and the loading l_j determines the slope of the probability curve.

In Section 4.1 we investigate descriptively whether the conditional proportion correct given total test score varied across OTL groups. In IRT language this is referred to as investigating item bias or using a more neutral term, differential item functioning. Standard IRT assumes invariate item functioning across different groups of individuals. A variety of bias detection schemes related to IRT have been discussed in the literature. Concerns about item bias due to instructional heterogeneity have recently been raised in the educational measurement literature. Conflicting results have been found in empirical studies. For example, Mehrens and Phillips (1986, 1987) found little differences in measurement characteristics of standardized tests due to varying curricula in schools, while Miller and Linn (1988), using the SIMS data, found large differences related to opportunity to learn, although these differences were not always interpretable. Muthén (1989) pointed out methodological problems in assessing differential item functioning when many items may be biased. He suggested a new approach based on a model which extends the standard IRT. The analysis is carried out by the LISCOMP program (Muthén, 1987). This approach is particularly suitable to the SIMS data situation with its item specific OTL information and it will be briefly reviewed here.

Let x be a vector of p OTL variables, one for each achievement item. The x variables may be continuous, but assume for simplicity that x_j is dichotomous with x_j = 0 for No OTL and x_j = 1 for OTL. Consider the modification of equation (2)

$$(4)\ y^* = l\,h + B\,x + e$$

where in general we restrict B to a diagonal p x p matrix. The diagonal element for item j is denoted b_j. The OTL variables are also seen as influencing the trait h,

$$(5)\ h = g'x + z$$

where g is a p-vector of regression parameter slopes and z is a residual.

It follows that

$$(6)\ P(y_j = i \mid h, x_j) = F[(-t_j + b_j x_j + l_j h)\, V(y^*_j \mid h)^{-\frac{1}{2}}]$$

In effect, then, the b_j coefficient indicates the added or reduced difficulty in the item due to OTL. Equivalently, using equation (4), we may see this effect as increasing y^*_j, the specific skill needed to solve item j.

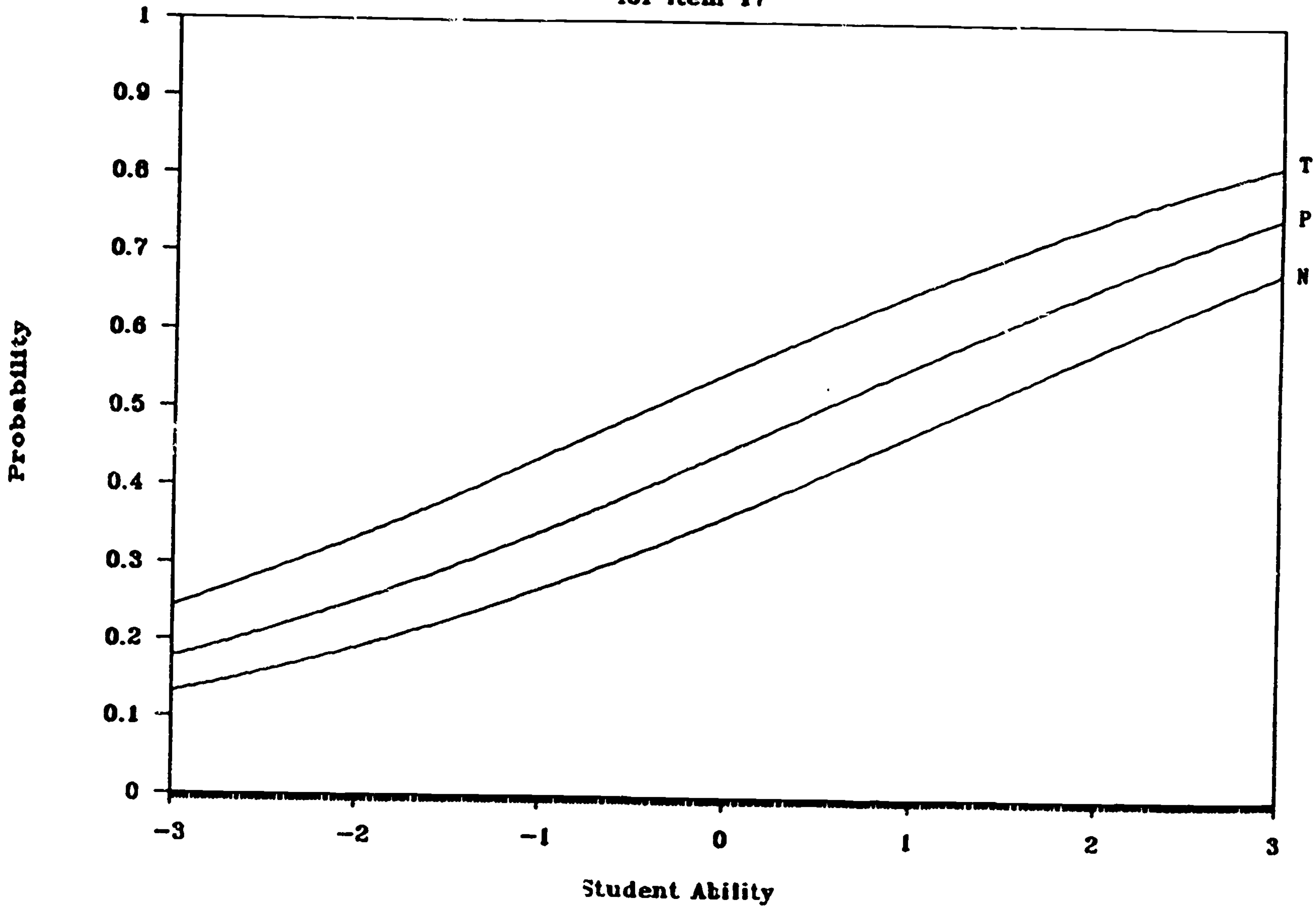
We note that this model allows for differential item functioning in terms of difficulty but not in terms of the slope related parameter l_j. This is in line with the data analysis findings of Section 4.1 where little difference in slopes of the conditional proportion correct curves was found across OTL groups (item 3 was an exception; we assume that this item will be reasonably well fitted by a varying difficulty model). More general modeling is in principle possible, but the data features do not seem to warrant such an extra effort.

This model disentangles the effects of OTL in an interesting way. Equation (5) states that OTL has an effect on the general achievement trait as measured by the g ccoefficients. Here we are interested in finding positive effects of instruction. Through the expected increase in h, such effects also have an indirect positive effect on the probability of a correct item response. The strength of h's effect on item j is measured by the coefficient l_j; (see equations (4) and (6)). In addition to the indirect effect of OTL for item j determined by g and l_j, there is also the possibility of a direct OTL effect on item j, which is determined by the b_j coefficient; see equations (4) and (6)). Any direct effect indicates that the specific skill needed to solve them, j, draws not only on the general achievement trait but also on OTL. The size of the g effect indicates the extent to which the achievement trait is sensitive to instruction. The size of the b_j effect indicates the amount of exposure sensitivity or instructional "over-sensitivity" in item j. While positive g effects correspond to a positive educational outcome, possible b_j effects are of less educational interest in that they demonstrate effects of teaching that influences very narrow content domains. From a text construction point of view items that show such exposure sensitivity are less suitable for inclusion in standardized tests, since they are prone to "item bias" in groups of examinees with varying instructional history. If such item bias goes undetected, IRT analysis is distorted. In the modeling presented here, however, exposure sensitivity is allowed for and the analysis does not suffer from the presence of such effects.

Muthén, Kao, and Burstein (1988) presents examples of analysis of exposure sensitivity using the dichotomous OTL groupings. However, we will first consider an

FIGURE 10

for Item 17



OTL, Prior OTL were used. Figure 10 shows the estimated item characteristic curves for item 17 having to do with acute angles. Since there are three OTL categories, there are three curves corresponding to three difficulty values.

Since the curves for both This Year OTL ar Prior OTL are above the No OTL curve, the b effects are positive for these two OTL groups. Exposure to the concept of acute angles produces a specific skill, which has the same effect as a reduced item difficulty, and this skill is not included in the general achievement trait. It is interesting to relate this finding to the percentage correct on item 17 broken down by OTL group as given in Table 1. Percentage correct increases d. matically from the No OTL category to the OTL categories, but the percentage correct is slightly higher for Prior OTL than for This Year OTL. For item 17 the Prior OTL students may do better than This Year OTL students, but Figure 10 shows that the recency of OTL gives an advantage for students at the same achievement trait level. Comparing the estimated item characteristic curves of Figure 10 with the empirical curves of Figure 5 we find a large degree of similarity but also differences. The estimated curves represent m re correct and precise estimates of these curves.

Muthén, Kao, and Burstein (1988) found substantial exposure sensitivity in items 3, 16, 17, 38, and 39, corresponding to solving for x, the product of negative integers, acute angles, percentages, and the coordinate system (see appendix). While items 3, 17 and 39 provided rather poor measurement of the achievement trait as indicated by their estimated l values, that was not the case for the other two. The authors hy, thesized that the exposure sensitivity corresponded to early learning of a definitional nature. Further analyses of the rotated form items, carried out by Kao (1989), supported this hypothesis. For example, the rotated forms showed exposure sensitivity for items covering square root problems. Overall, about 15 30% of the items exhibit mild exposure sensitivity, while only about 10 - 15% exhibit strong exposure sensitivity. We may note that these percentages are considerably lower than the Miller and Linn (1988) findings using related parts of the SIMS data and standard IRT methodology. The effects of OTL on the achievement trait will be discussed in later sections.

6. Multidimensional modeling

Standard IRT modeling assumes a unidimensional trait as was also done in the previous section. For a carefully selected set of test items, this is often a good approximation. However, in many achievement applications, it is reasonable to assume that sets of items draw on more than one achievement trait.

Muthén (1978) presented a method for the factor analysis of dichotomous items, where the model is

(7) $y^* = L\ h + e$

(8) $V(y^*) = L\, y\, L' + Q$

where L is a p x m factor loading matrix, y is a factor covariance matrix, and Q is a diagonal matrix of residual variances. In line with item analysis tradition (see Lord & Novick, 1968), Muthén fitted the model to a matrix of sample tetrachorics. For an overview of factor analysis with dichotomous items, see Mislevy (1986).

Although of great substantive interest, models with many minor factors are very hard to identify by usual means of analysis. For instance, assume as we will for the SIMS data, that a general achievement factor is the dominant factor in that it influences the responses to all items. Assume that, in addition to this general factor there are several specific factors, orthogonal to the general factor, that influence small sets of items of common, narrow content. It is well known that such models with continuous data cannot be easily recovered by ordinary exploratory factor analysis techniques involving rotations. This problem carries over directly to dimensionality analysis of dichotomous items using tetrachoric correlations.

Consider as an illustration of the problem an artificial model for forty dichotomous items. Assume that one general factor influences all items and eight specific factors each influence a set of five items. Let the general factor loadings be 0.5 and 0.6 while the specific factor loadings are 0.3 and 0.4. Let the factors be standardized to unit variances and let the factors be uncorrelated. The eigenvalues of the corresponding artificial correlation matrix are shown in Figure 11. Such a "scree plot" is used for determining the number of factors in an item set. The number of factors is taken to correspond to the first break point in the plot where the eigenvalues level off. If the first eigenvalue is considerably larger than the others and the others are approximately equal, this is usually taken as a strong indication of unidimensionality. Figure 11 clearly indicates unidimensionality despite the existence of the eight specific factors. There would be no reason to consider solutions of higher dimensionality.

As a comparison, Figure 12 shows the eigenvalues for the tetrachoric correlation matrix for the 39 core items of the SIMS data. The two eigenvalue plots are rather similar. Models similar to the artificial one considered above have been studied by Schmid and Leiman (1957), where it was pointed out that the above hypothesized nine-factor model can also be represented as an eight-factor model with correlated factors.

FIGURE 11

Scree Plot for Tetrachoric Correlations
with Artificial Model for 40 Items

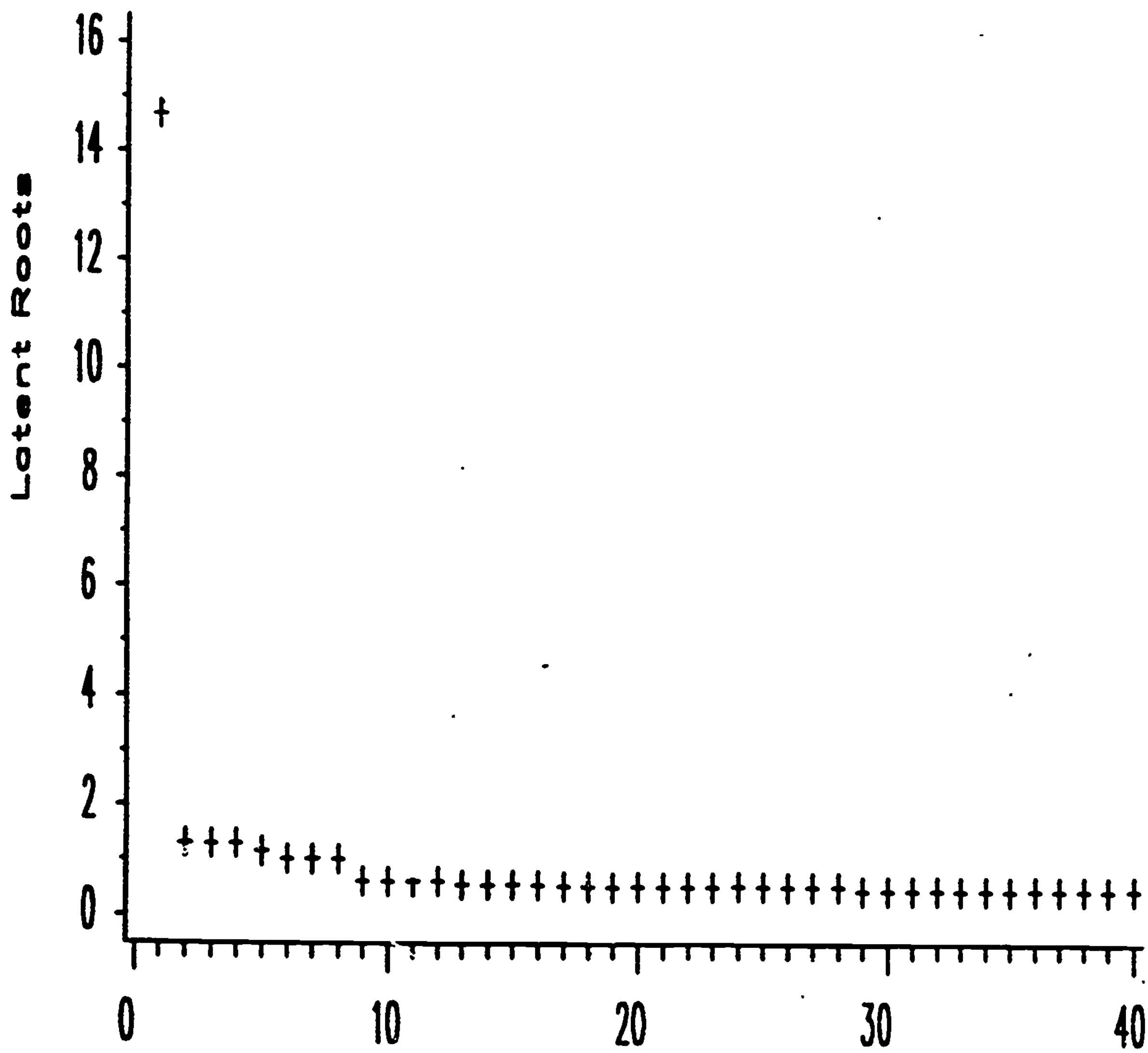
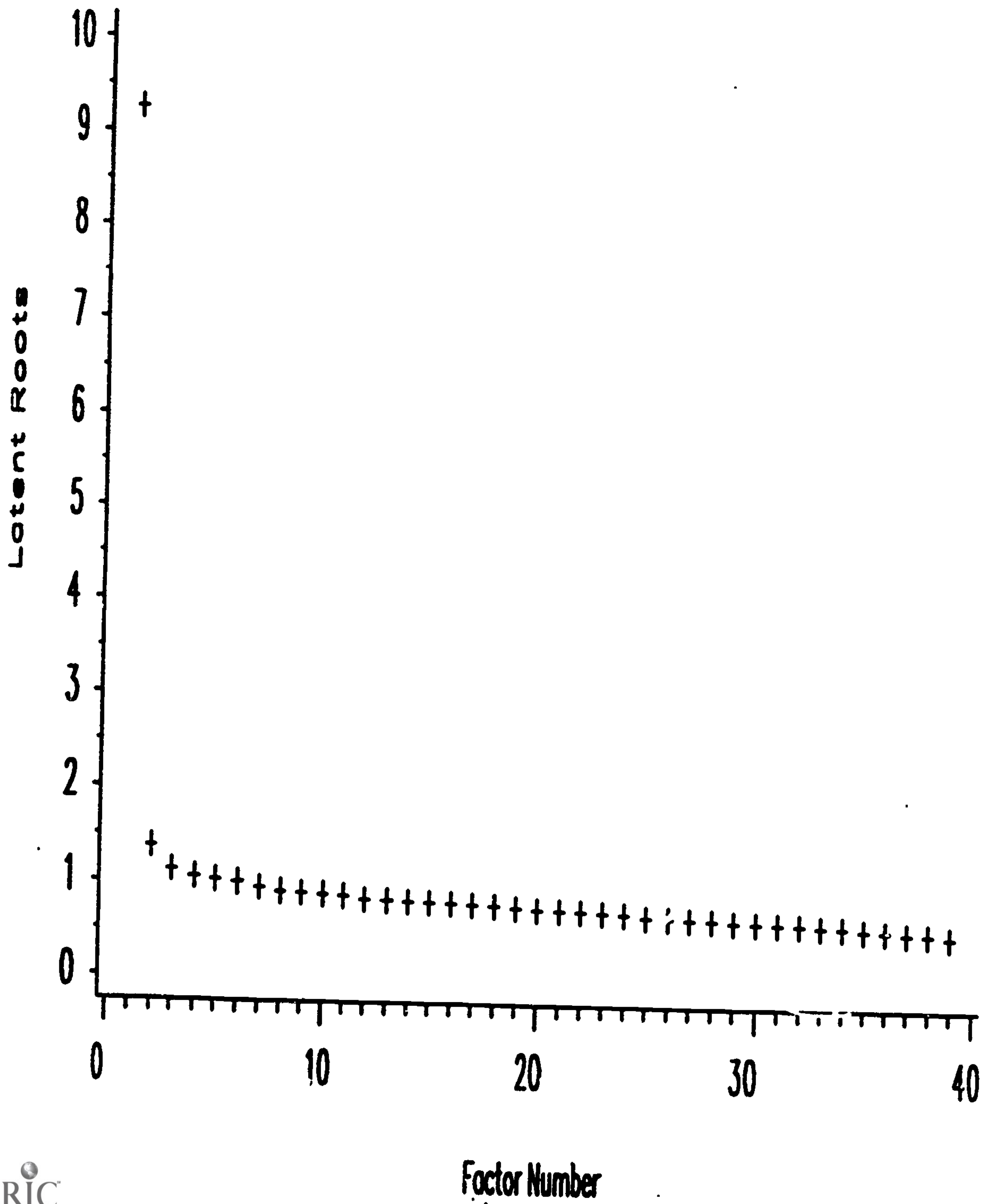


FIGURE 12

Scree Plot of Latent Roots for 39 Items Based on Tetrachorics



Each of the eight factors may be viewed as a function of both a general, second-order factor and the corresponding specific factor of the nine-factor model. The specific factor is then viewed as a residual contribution, orthogonal to the second order factor. Hence, Schmid and Lieman used the term hierarchical factor analysis. Using exploratory factor analysis on the artificial correlation matrix, an oblique rotation of the eight factor solution did indeed identify the eight correlated factors of such a hierarchical reformulation to the model. Schmid and Lieman (1957) gave formulas for transforming such a solution back to the original model with a general factor and eight specific factors, all factors being uncorrelated. However, without knowing the correct number of factors, there would have been no guide to choosing this eight-factor solution.

The usefulness of hierarchical factor analysis has recently been pointed out by Gustafsson (1988a, b). He proposed to circumvent the difficulties of using exploratory factor analysis by formulating confirmatory factor analysis models. Hypothesizing a certain specific factor structure in addition to a general factor, the confirmatory model enables the estimation of factors with very narrow content. Applications of this type of modeling to the SIMS data are being considered by the author in collaboration with Burstein, Gustafsson, Webb, Kim, Novak, and Short. In line with our previous modeling, we may write a simple version of this model as

(9) $y^*_j = l_{G_j} h_G + l_{S_j} h_{S_k} + e_j$

where y^* is the latent response variable for item j (cf. the Section 4 model), h_G is the general achievement factor, h_{S_k} is the specific factor for item j, and e_j is a residual. The three right hand side variables are taken to be uncorrelated. This means that the items belonging to a certain specific factor correlate not only due to the general factor but also due to this specific factor.

In this simplified version of the model, it assumed that each item measures only one specific factor. For identification purposes we assume that each specific factor h_{S_k} is measured by at least two items. Also for identification purposes, our baseline model will set $l_{S_j} = 1$ for all j's, although this can be relaxed as a need arises as well be discussed below. In this way, the general factor is assumed to influence each item to a different degree, while the specific factor has the same influence on all items in the corresponding set.

The multidimensional confirmatory factor analysis model allows an interesting variance component model interpretation. Standardizing the general factor variance to unity, while letting the specific factor variances be free parameters, the model implies a decomposition of the latent response variable variances into a general factor component, a specific factor component, and an error component:

$$(10)\quad V(y^*_j) = lG_j^{\,2} + YS_k + q_j$$

where YSk is the variance of the specific factor k. Since the items are dichotomous, the variances of the y*'s are standardized to one by restrictions on the qj's. The relative sizes of the first two terms on the right hand side of (10), the general and the specific components, are of particular interest. The specific component can also be interpreted as the average correlation remaining between items belonging to specific factor k when holding the general factor constant. The model can be estimated by confirmatory factor analysis techniques for dichotomous items using the LISCOMP computer program, see Muthén (1978, 1987).

The SIMS items of the core and the rotated forms were classified into subsets corresponding to specific factors defined both by content and procedure. Examples of the narrow item domains that were considered are: Arithmetic with signed numbers (core items 3, 16, 25), percent calculations (core items 2, 34, 36, 38), estimation skills (size, distance; core items 6, 8, 9), and angular measurements (core items 17, 19, 21, 22).

The analysis steps are as follows. For a given hypothesized set of specific factors, a confirmatory factor analysis run can be performed. The initial model may then be refined in several steps. An inappropriate combination of items for a specific factor gives rise to a low or negative variance component estimate for this specific factor. Modifications may be assisted by inspection of model misfit indices. For this model a useful index is related to the loadings of the specific factors, lS_j, which are fixed to unity in the baseline model. The sign and size of the derivatives of these loadings are of interest. A positive value for a certain item indicates that if the loading is free to be estimated, the estimated value will be smaller than one. In effect, this allows the estimate of the variance component for the specific factor at hand to increase. This is because the specific variance component is related to the average correlation of the specific factor items, conditional on the general factor, where the decrease in the factor loading for a certain item means that the contribution from this item is weighted down. Thus modifying the initial analysis, items that obtain very low or negative specific factor loadings are candidates for exclusion from the set assigned to this specific factor. This modification process may be performed in several iterations. In the analyses performed for the SIMS data, this procedure appeared to produce substantively meaningful results in that the items that were singled out clearly had features that distinguished them from the others in the set.

Table 2 gives the estimated variance components for core items corresponding to three of the specific factors.

Table 2
Variance Components for Selected Items from the Core*

Item	General Factor	Specific Factors: Percent	Estimate	Angular Measurement
AR02	33(24)	9(9)		
AR34	39(32)	9(9)		
AR36	32(27)	9(9)		
AR39	35(26)	9(9)		
ME06	20(14		9(10)	
ME08	38(27		9(10)	
ME09	38(29)		9(10)	
GE17	28(17)			11(12)
GE19	17(12)			11(12)
GE21	24(17)			11(12)
GE22	43(30			11(12)

*Given in parenthesis is the estimate when controlling for mean level heterogeneity. (See section7)

It is seen that the variance contribution from the specific factors can be as large as 50% of that of the general factor and are therefore of great practical significance. This is particularly so since the sets of items for a specific factor correspond closely to instructional units. Analyses of the rotated forms replicated most of the specific factors found for the core.

The confirmatory factor analysis procedure described is a cumbersome one involving many iterations and many subjective decisions. An attempt was therefore made to find an approach which would involve fewer steps and a more objective analysis. It was reasoned that if the influence of the general factor could be removed from the item correlations, the remaining correlations would be due to the specific factors alone. Such residual correlations could then be factor analyzed by regular exploratory techniques, at least if nesting of specific factors within each other was ignored. Given a proxy for the general factor, the residual correlations could be obtained by bivariate probit regressions of all pairs of items on the proxy, using the LISCOMP program.

An attempt was first made to approximate the general factor for the posttest core items with the posttest total score. However, this produced almost zero residual correlations. Instead, the pretest total score was used for the posttest items. An exploratory factor analysis of these residual correlations, using an orthogonal rotation by Varimax, resulted in eleven factors with eigenvalues greater than one. The interpretation of these factors showed an extraordinary high degree of agreement with the specific factors previously obtained. The best agreement was obtained for factors that had obtained the largest variance component estimates. The exploratory analysis also suggested a few items to be added to the specific factors as defined earlier. The agreement of these two very different approaches is remarkable and it is interesting that the pretest score appears to be a better proxy for the general factor at the posttest occasion than the posttest score. This may indicate that the general factor is a relatively stable trait related to the achievement level before eighth grade instruction; we note from Table 1 that This Year OTL is the most prevalent category. Controlling for posttest score may in contrast control for a combination of the general factor and specific factors.

It is interesting to note that analyses of the core items administered at the pretest gave very similar results in terms of specific factors identified by the confirmatory approach. This indicates stability of the specific factors over the eighth grade. Attempting to compute residual correlations for exploratory factor analysis again gave near zero values when controlling for the total score, the pretest this case, and this approach had to be abandoned.

7. Modeling with heterogeneity in levels

The factor analysis of the previous section was performed under the regular assumption of identically distributed observations, that is all students are assumed to be sampled from the same population with one set of parameters. However, we have already noted that the students have widely varying instructional histories and that the homogeneity of student populations is not a realistic assumption. This is a common problem in educational data analysis which has been given rather little attention. We may ask how this heterogeneity affects our analysis and if it can be taken into account in our modeling.

Muthén (1988a) considers covariance structure modeling in populations with heterogeneous mean levels. This research considers both the effect of incorrectly ignoring the heterogeneity and proposes a method to build the heterogeniety into the model. The method is directly applicable to the multidimensional factor analysis model considered in the previous section and can also be carried out within the LISCOMP framework. Consider the model of equation (7)

(11) $y^* = L\,h + e$

In the previous section we made the usual standardization of E (hi) = O for all observations i and assumed V (h_i) = y. However, we know that it is unrealistic to assume that for example students from different class types have the same factor means levels and we may instead want to assume that the means vary with class type such that for student i in class c we have E (h_{ic}) = a_c. As pointed out in Muthen (1988a) this may be accomplished by considering in addition to (11) the equations

(12) $h_{ic} = G\,x_c + z_{ic}$

where x_c represents a vector of class type dummy variable values for class c, G is a parameter matrix, and z_{ic} is a residual vector for student i in class c. We assume that conditional on class type membership the factor means vary while the factor covariance matrix remains constant,

(13) $E\,(h_{ic} \mid x_c) = Gx_c$

(14) $V\,(h_{ic} \mid x_c) = Y$

The modeling also assumes that the matrices L and Q are constant across class types, so that

(15) $E\,(y^* \mid x_c) = L\,G\,x_c$

(16) $V\,(y^* \mid x_c) = L\,Y\,L' + Q$

It is interesting to note that the assumption of constancy of the conditional covariance matrix V (y^* I x_c) is in line with the findings of constancy of the homogeneity of correlations found in Section 4.2.

The structure imposed on the parameter matrices of (15) and (16) may correspond to an exploratory or a confirmatory factor analysis model. Muthen (1988a) points out that the conditional covariance matrix of (16) is not in general the same as the marginal covariance matrix V (y^*). In our context this means that even when we have the same factor analysis structure in the different class types this covariance structure does not hold in the total group of students. The approach outlined here essentially provides a mean-adjusted analysis of pooled covariance matrices assumed to be equal in the population. In our situation the analysis effectively is carried out on pooled tetrachoric correlation matrices. This modeling has two important outcomes. The dimensionality analysis can be carried out without distortion due to the differences in factor mean levels across class types and the factor mean levels can be estimated.

The above mean-adjusted analysis was carried out on the SIMS core items using the multidimensional factor model from Table 2 of the previous section. Factor mean differences were allowed for class type using three dummy variables and also gender. We will concentrate our discussion of the results on the factor structure. Despite large mean differences across class type for the general achievement factor, a factor structure very similar to the previous one emerged. The same specific factors showed large and small variances, respectively. Hence, the potential for a distorted structure is not realized in these data. The results are presented in parentheses in Table 2. It is seen that the variance contributions to the general factor are considerably reduced as compared to the first approach.

The reduction in variance contribution from the general factor is natural since holding class type constant reduces the individual differences in the general achievement trait due to selection of students. If the inference is to the mix of students encountered in the SIMS data the unreduced variation in the trait is the correct one, but this variation is not representative for a student from any given class type. It is also interesting to note that the specific factor variances are not similarly reduced by holding class type constant, presumably indicating that these specific skills are largely unrelated to the student differences represented by class type.

8. Estimation of trait scores

Sections 5, 6, and 7 have considered various factor analysis models for the achievement responses. Assuming known or well-estimated parameter values for these models it is of interest to estimate each student's score on factors of these models. For the standard, unidimensional IRT model estimation of the trait values is a standard task which may be carried out by maximum likelihood, Bayes' modal (maximum a posteriori), or expected a posteriori estimator (see for example Bock & Mislevy, 1986). The instructionally sensitive models we have considered for the SIMS data have however brought us outside this standard situation in the following three respects:

(i) In line with Section 5 we want to consider factor score estimation that takes into account that certain items have different difficulty level depending on the students' OTL level.

(ii) In line with Section 6 we want to consider factor scores for both the general achievement factor and the specific factors in the multidimensional model.

(iii) In line with Section 7 we want to consider factor scores estimation that takes into account differences in student achievement level.

We note that (i) and (iii) are quite controversial since these points raise the issue of estimating achievement scores based not only on the student's test responses but also his/her instructional background. For example Bock (1972) has argued that prior information on groups should not be used in comparisons of individuals across groups. Nevertheless, it would seem that students who have had very limited OTL on a set of test items will be unfairly disadvantaged in comparison with students with different instructional exposure. The aim may instead be to obtain achievement scores for given instructional experiences.

Point (ii) is of considerable interest. While a rough proxy for the general achievement score is easily obtainable as the total test score, the adding of items corresponding to specific factors would involve only a few items resulting in a very unreliable score. As a contrast, estimating the specific factor scores draws on the correlated responses from all other items.

The following estimation procedure was discussed in Muthen and Short (1988) and handles all three cases above. For various density and probability functions, g, consider the posteriori distribution of the factors of h,

(14) $g(h \mid y, x) = f(h \mid x)\, g(y \mid h, x) / g(y \mid x)$

Here, the first term on the right hand side represents a normal prior distribution for h conditional on x, where as before x represents instructional background variables such as OTL and class type. In line with Section 7 the factor covariance matrix may be taken as constant given x, while the factor means may vary with x. The second term on the right hand side represents the product of the item characteristic curves, which may vary in difficulty across OTL levels as discussed in Section 5.

Muthén and Short (1988) considered an example of the situation of (i) and (iii). They generated a random sample of 1,000 observations from a model with forty items measuring a unidimensional trait. Observations were also generated from forty OTL variables and five other background variables. All background variables were assumed to influence the trait while the first twenty OTL variables had direct effects on their corresponding items, giving rise to exposure sensitivity in these items. Among other results, Muthén and Short considered differences in factor score estimates using the above method and the traditional IRT method. In Table 3 comparisons of the two corresponding score distributions are presented by quartiles, broken down in two parts - students with a high total sum of OTL and students with a low sum. The table demonstrates that for students of the low OTL group, estimated scores are on the whole higher with the new method, corresponding to an adjustment for having had less exposure, while for the high OTL group the estimated scores are on the whole lower for the new method.

Ongoing work by Muthén and Short investigates situation (ii) and the precision with which scores for specific factors can be estimated. Once the estimated factor scores have been calculated they may conveniently be related to various instructional variables and may also be studied for change from pretest to posttest.

Table 3
Trait Estimates by Traditional and New Approaches*

LOW OTL GROUP					
	TRADITIONAL				
NEW	25%	50%	75%	100%	TOTAL
25%	136 -1.323 -1.255	6 -0.610 -0.724	0	0	142 -1.293 -1.233
50%	10 -0.783 -0.624	125 -0.361 -0.338	5 0.037 -0.119	0	140 -0.375 -0.351
75%	0	13 -0.094 0.058	111 0.309 0.316	7 0.827 0.691	131 0.297 0.311
100%	0	0	6 0.691 0.834	124 1.282 1.308	130 1.255 1.286
TOTAL	146 -1.286 -1.212	144 -0.347 -0.318	122 0.317 0.324	131 1.257 1.275	543

Table 3 (Continued)
Trait Estimates by Traditional and New Approaches*

HIGH OTL GROUP

NEW	TRADITIONAL 25%	50%	75%	100%	TOTAL
	99	9	0	0	108
25%	-1.306	-0.578			-1.245
	-1.349	-0.743			-1.298
	5	94	12	0	111
50%	-0.726	-0.340	0.049		-0.315
	-0.581	-0.366	-0.119		-0.349
	0	3	110	5	118
75%		-0.167	0.345	0.870	0.355
		-0.222	0.322	0.640	0.327
	0	0	6	114	120
100%			0.653	1.386	1.349
			0.782	1.334	1.306
TOTAL	104	106	128	119	457
	-1.278	-0.355	0.332	1.364	
	-1.312	-0.389	0.302	1.305	

*Entries are:
Frequency
mean value by the traditional approach
mean value by the new approach

9. Predicting achievement

Given the explorations of the previous sections, we may attempt to formulate a more comprehensive model for the data. Muthén (1988b) proposed the use of structural equation modeling for this task. He discussed a model which extends ordinary structural modeling to dichotomous response variables while at the same time extending ordinary IRT to include predictors of the trait. He studied part of the SIMS data using a model which attempted to predict a unidimensional algebra trait at the posttest occasion using a set of instructional and student background variables from the pretest. The set of predictors used and their standardized effects are given in Table 4. While pretest scores have strong expected effects, class type, being female, father being in the high occupational category, and finding mathematics useful to future needs also had strong effects. The OTL variables had very small effects overall, perhaps due to the fact that each item's OTL variable has rather little power in predicting this general trait.

Table 4
Structural Parameters with the latent Construct as Dependent Variable

Regressor	Estimate	Estimate/S.E.
PREALG	0.68	11
PREMEAS	0.45	7
PREGEOM	0.33	5
PREARITH	2.09	16
FAED	0.07	1
MOED	0.02	0
MORED	0.18	3
USEFUL	0.45	7
ATTRACT	0.04	1
NONWHITE	-0.02	0
REMEDIAL	0.07	1
ENRICHED	0.22	3
ALGEBRA	0.56	4
FEMALE	0.14	[illegible]
LOWOCC	0.02	1
HIGHOCC	0.12	3
MISSOCC	0.05	2
NONWXREM	0.10	1
NONWXENR	0.19	3
NONWXALG	-0.18	-1
PREARITH X REM	-1.45	-3
PREARITH X ENR	-0.10	-1
PREARITH X ALG	-0.54	-2
NONW X PREARITH	-0.19	-1

Given the analysis results of the previous sections, this modeling approach can be extended to include a multidimensional model for both the set of pretest and posttest items, predicting posttest factors from pretest factors, using instructional and student background variables as covariates, and allowing for differential item functioning in terms of exposure sensitivity. This work is in progress.

10 Analyzing change

The structural modeling discussed in the previous section is also suitable for modeling of change from pretest to posttest. In Section 4.3 we pointed out that in terms of change the SIMS data again exemplified complex population heterogeneity. For each item a student may belong to either of three OTL groups, corresponding to two types of no new learning and learning during the year. To again reach the goal of instructionally sensitive psychometrics as stated in Section 3 for this new situation, we should explicitly model this heterogeneity. However, to properly model such complex heterogeneity is a very challenging task and this work has merely begun.

A basic assumption is that change is different for groups of students of different class types and OTL patterns. In a structural model where posttest factors are regressed on pretest factors the slopes may be viewed as varying across such student groups, where students groups for which a large degree of learning during the year has taken place, as measured by the set of OTL variables, are assumed to have steeper slopes than the other students. This methods area shows a very large degree of scarcity of psychometric work.

References

Bock, R. D. Review of The dependability of behavioral measurements by Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. Science, 1972, 178, 1275-12275A.

Bock, R. D., & Mislevy, R. J. (1986). Adaptive EAP estimation of ability in a microcomputer environment. Applied Psychological Measurement, 6, 431-444.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O.., McKnight, C. C., & Cooney, T. J. (1985). Second International Mathematics Study: Summary report for the United States. Champaign, IL: Stipes.

Gustafsson, J. E. (1988a). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, Vol. 4. Hillsdale, NJ: Lawrence Erlbaum.

Gustafsson. J. E. (1988b). Broad and narrow abilities in research on learning and instruction. Learning and individual differences: Abilities, motivation, methodology. Symposium conducted at the Minnesota Symposium, Minnesota.

Kao, C. F. (1989). An investigation of instructional sensitivity in mathematics achievement test items for U.S. eighth grade students. Doctoral dissertation, University of California, Los Angeles, Los Angeles. In progress.

Lord, F. M. (1980). Application of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. Journal of Educational Measurement, 23, 185-196.

Mehrens, W. A., & Phillips, S. E. (1987). Sensitivity of item difficulties to curricular validity. Journal of Educational Measurement, 24, 357-370.

Miller, M. D., & Linn, R. L. (1988). Invariance of item characteristic functions with variations in instructional coverage. Journal of Educational Measurement, 25, 205-219.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

Muthén, B. (1988a). Covariance structure modeling in heterogeneous populations: Mean-adjusted analysis. Los Angeles: University of California, in preparation.

Muthén, B. (1988b). Some uses of structural equation modeling in validity studies: Extended IRT to external variables. In H. Wainer & H. I. Pruan (Eds.), Test Validity. Hillsdale, NJ: Erlbaum, 213-238.

Muthén, B. (1989). Using item-specific instructional information in achieving modeling. Psychometrika, in press.

Muthén, B., Kao, C-F., & Burstein, L. (1988). Instructional sensitivity in mathematics achievement test items: Application of a new IRT-based detection technique. Paper presented at 1988 American Educational Research Association meeting. New Orleans.

Muthén, B., & Short, L. M. (1988). Estimation of ability by IRT models allowing for heterogeneous instructional background. Paper presented at 1988 American Educational Research Association meeting. New Orleans.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. Psychometrika. 22, 53-61.

NON-COGNITIVE DATA
A Cross-national Perspective

W. Todd Rogers
University of British Columbia

I welcome this opportunity to share with you my experience when analyzing the non-cognitive attitudinal data collected as part of SIMS. The results of that analysis may be found in the attached paper and from which I will speak:

Rogers, W. T., & O'Shea, T. (1985). A comparative analysis of attitudes toward mathematics of senior high school students in British Columbia, Ontario, and the United States. Canadian and International Education/Education Canadienne et Internationale, 14, 39-58.

After summarizing this study and its results, I would like to make some additional remarks concerning my view of secondary analysis and what can be done to encourage such analyses.

Secondary Analyses of Previously Collected Data: Some Comments

Studies such as SIMS, The National Assessment of Educational Progress, and state and provincial (in the case of Canada) assessments offer a rich source of data for primary and secondary data analyses. This is particularly so in light of the periodic replication or repetition of these studies.

Invariably, the first focus of these studies is to answer the questions initially posed when seeking funding support. But, in recognition of the massive data sets required to answer these primary questions, and the relative limited amount of available funds, the principal investigators often include in their rationale for seeking support for the study that the data will be made available to other researchers for secondary analyses. The supposition here is that the data collected are amenable to addressing questions other than those included in the primary set and, therefore, costs can be amortized across a wider base or set of studies. In my opinion, this is appropriate. Indeed, I advocated such an approach when assisting the Ministry of Education in British Columbia establish its provincial assessment program.

The paper I am discussing can be classified as a secondary study of the data from the SIMS. Neither Tom O'Shea or I were involved in the creation of the data collection instruments. I did have some early involvement in the sample design for the province. Therefore, I think we can be classified as people who accepted the invitation to look at the IEA data from the Second Study. Our research was supported by the Social Sciences and Humanities Research Council of Canada as part of a larger research project (under the direction of Dr. David Robitaille) designed to complete further analyses of the IEA data beyond the analyses included in the initial IEA proposal.

What was our experience? First I would like to thank the SIMS project directors for making available the data we used. Further, as questions arose concerning the development and validation of the attitude scales or the nature of the data set, particularly with respect to sampling weights, they were graciously and quickly answered. Such initial and continued cooperation is crucial to the success of a secondary study. I therefore recommend that:

(i) support for such cooperation be included in the initial funds provided for the initial or primary study.

One problem which we had, and which we felt we could do something about, was how to treat missing data. To our surprise, the file we received had not yet been edited for missing data. Our concern was how to treat such data so that our treatment was consistent with that employed by the IEA in its own analyses, and in other analyses of the IEA data. In our study, students who omitted more than three quarters of the items of a score were removed from the file; missing data on individual items for an individual student were assigned the mid-point value of three. But is this what others would have done? A third and more difficult problem to solve centered on matching student class data with teacher data in schools from which more than one class was selected. This problem likely arose during data collection, when the data were initially collected, or in data entry, when the data were transferred to the computer data file. What ever the source, it was a problem which we were not able to solve. Consequently, we removed unmatched classes and teachers. To ensure comparability of data sets across different analyses of the same data set, I recommend that

(ii) data files be edited by the primary investigators to their (the data) release to those wishing to do additional analyses of the data, and the editing procedure used be clearly described in supporting documentation.

A related recommendation is that

(iii) an intermediate data file containing descriptive statistics be provided by the primary investigators to act as a check against which the secondary researchers may verify that they have correctly accessed the data file.

The last problem with which we tussled revolved around the measurement scales. Clearly, the scales employed by SIMS were developed prior to our involvement. Thus, we had to use what was available. As might be expected, we would have liked to have made some changes. However, unlike the concerns above, we were not able to. Indeed, fundamental to answering a research question is the validity of the data collection instrument--tests, survey forms, interview schedules--in terms of that question. If a researcher perceives that the instruments used and, hence, the data collected are not appropriate for the research questions posed, then it is unlikely that he/she will request a copy of the data set. I deliberately used the work perceived, for very often the development and validation of the instruments used in the primary study are not adequately described for others to gain a full understanding of these instruments and their use. We found it necessary, for example, to contact the IEA officials on more than one occasion for information beyond that contained in the documentation provided. It is therefore recommended that

(iv) complete documentation in a form similar to that called for in the Standards for Educational and Psychological Tests be provided by the primary investigators.

Consideration of the above issues together reveals that many of the concerns are traceable to documentation. Recognition must be given to the needs of a secondary researcher so that he/she comes to feel ownership of the data provided in much the same way as ownership is felt when a researcher collects his/her own data. No small feat, provision should be made when seeking support for the initial study to include proper and full documentation of all elements of the primary research for use in a secondary analysis of that data.

A Comparative Analysis of Attitudes Toward Mathematics of Senior High School Students in British Columbia, Ontario, and the United States

Todd W. Rogers and Thomas O'Shea

The International Association for the Evaluation of Educational Achievement (IEA) recently conducted its second study of mathematics. The portion of the study reported here deal with the attitudes toward mathematics of Grade 12 students in British Columbia, Grade 12 and 13 students in Ontario, and Grade 12 students in the U.S.A., some of whom were enrolled in Calculus courses.

Students responded to Likert-type items making up the following scales: Mathematics in School, Calculators and Computers, Home Support for Mathematics, Mathematics and Me. Mathematics and Utility, Mathematics and Gender, and Mathematics as a Process. For each scale a 5 x k ("country"-by-item, k the number of items) median polish was used to analyze the unique and joint effects of country and item. Results are reported in three areas: students' opinion on the mathematics curriculum, personal perceptions of mathematics, and views on the discipline of mathematics.

In general, students reported remarkably similar views on all scales. We attribute this to the pervasive influence of American educational theory and practice and to the structured nature of the mathematics curriculum. Students also appeared to value mathematics for its practicality rather than for its intrinsic worth.

Introduction

In keeping with the basic design of the comparative assessments conducted by the International Association for the Evaluation of Educational Achievement (IEA), the Second IEA Study of Mathematics included an assessment of student opinions, preferences, and attitudes toward a number of aspects of mathematics and mathematics education. IEA is an association of educational research organizations and ministries of education whose primary goals are to conduct educational research on an international level and to assist member-states in undertaking cooperative research projects. IEA has conducted international surveys in the past, including the First Mathematics Study (Husen, 1967) and the Six Subject Survey (Peaker, 1975; Walker, 1976). In the Second Study of Mathematics the attitudinal topics ranged from the nature of mathematics and its role in society to specifics of the mathematics curriculum. As well, attitudes of teachers toward mathematics as a process were assessed.

Justification for assessing attitudes comes not only from the tradition of IEA studies, but also from the importance attached to affective variables in research and in the assessment of outcomes and processes of schooling. There continues the search to find predictors from the affective domain, as well as from among personalogical and process variables, to increase the accuracy of statistical models to explain and predict variation in mathematics achievement. Alternatively, affective measures are thought to reflect outcomes of schooling. Affective variables, then, become outcomes to be explained or predicted rather than variables to be used to explain or predict.

More relevant to the present study, however, is the use of affective measures to assess how students perceive and respond to what is actually happening in schools. Student responses to the affective items included in this survey reflect from the point of view of the learner what is occurring in the mathematics classroom. Results on the survey items not only are indicative of what happens in classrooms, but also reflect prevailing opinions about mathematics in broader social contexts. Except for the influence of school, there would be little reason for a student to have formed an opinion about the importance of a particular mathematics concept or about the usefulness of learning that concept.

Of particular interest in this paper are the differences and similarities between the opinions, preferences, and attitudes of students from five "countries"--Grade 12 in British Columbia (B.C.) Canada; Grade 12 and Grade 13 in Ontario (Ont.), Canada; and Pre-Calculus and Calculus in the United States (U.S.). Factors which suggest that the responses would be more similar than different include the close proximity of Canada and the United States, the pervasive influence of American educational theory and practice (Andrews & Rogers, 1982), and the structured nature of the mathematics curriculum. Similarities should be particularly evident for students at comparable levels of education: Grade 12 in B.C., Grade 12 in Ont., and Pre-Calculus in the U.S.; and Grade 13 in Ont. and Calculus in the U.S. On the other hand, Canada's commitment to maintaining and encouraging its own identity, particularly in Ontario where all textbooks must be authored by Canadians, may result in differences in opinions and attitudes.

In addition to the student responses, the views of teachers concerning the nature of mathematics as a discipline were identified. Teachers also responded to four items related to the second mathematics curriculum. Three of these were involved in the set of 15 presented to the students. The absence of a rationale supporting the selection of these three items and the use of only four raises serious questions about the comprehensiveness of coverage. For this reason, the teacher responses to these items were not included.

Structure of the Items and Scales

Because there existed no apparent consensus in mathematics education of what should be measured in the affective domain, the International Mathematics Committee (IMC) proposed the following four guidelines for constructing attitude items:

- Items should address issues of importance to mathematics educators,
- Responses to items should provide useful descriptive information,
- Items should permit the formation of scales, and
- There should be items from the first IEA study of mathematics (Kifer, 1979).

Based on these guidelines, and following general discussions involving IMC members and representatives of the IEA General Assembly, seven general domains were identified. Table 1 contains a short description of each domain and the final number of items in each. A copy of the final form of each scale is provided in the Appendix of this paper.

Briefly, initial items were selected from the first IEA mathematics survey, the National Assessment of Educational Progress in the U.S., and from other existing mathematics attitude scales. New items were written to provide adequate size pools for each domain. Although responses to the items were structured differently depending on the scale, a common five-point Likert format was adopted. An example of an item taken from Mathematics in School and illustrative of the differences in structure is as follows:

Solving Equations

a) Important - Not important

Very Important	Important	Undecided	Not Important	Not at all Important

b) Easy - Hard

Very Easy	Easy	Undecided	Hard	Very Hard

c) Like - Dislike

Like a Lot	Like	Undecided	Dislike	Dislike a Lot

These item pools were pilot tested in the U.S., and the results were used to select appropriate items and to form scales. The scales were then field tested in international trials to evaluate the extent to which items translated well, were acceptable to the participating countries, and possessed desirable psychometric properties. The content validity of the scales was reported to be satisfactory by mathematics educators in the participating countries. Estimates of internal consistency derived from the field testing, however, revealed that analyses of the item level, and not at the scale level, would be appropriate for Calculators and Computers and Mathematics as a Process (Kifer, 1979, personal communication).

Samples

Since the interest in the Second IEA Study of Mathematics was focussed on teaching and learning at the class level, probability samples of classes were selected from each population of interest. A basic sample design was recommended, but countries were permitted to make approved modifications. Table 2 summarizes the stratification and selection procedures employed for each population. As shown, each sample may be described as a deeply stratified, multi-stage probability sample.

The overall response rates at the class levels were 90% for B.C. and 86% for Ont. For the U.S., school districts, schools, and classes were oversampled to allow for refusals. The cooperation rates at each stage were approximately 50%, 75%, and 90%. Despite these lower values, the desired sample sizes were achieved (Garden, 1985). Furthermore, the obtained samples were comparable to other U.S. samples (Garden, 1985, Appendix 3).

Data Analysis

To facilitate examination of the relationship between student and teachers responses on the Mathematics as a Process items, the item data files were first edited to remove respondents for whom data were missing on entire scales, and then to remove unmatched classes and teachers. Individual items containing missing data were assigned the mid-point value three (undecided). At the same time, the polarity of negatively

Table 1
Attitude Items and Scales

Items	Domain	Description of Items	No of Scales
1.	Mathematics in School (a)	Attitudes toward mathematical topics and activities believed to be universally part of mathematics curricula. Three dimensions were considered: importance, difficulty, liking.	15
2.	Calculators and Computers	View of the nature and usefulness of hand calculators and computers.	8
3.	Home Support for Mathematics	Parental ability and support for the study of mathematics.	9
4.	Mathematics and Me	Personal reaction to the study of mathematics in terms of feelings, enjoyment, competence, anxiety, and willingness to study more mathematics.	15
5.	Mathematics and Utility	View of the practical value of mathematics in preparing for an occupation and in everyday life.	8
6.	Mathematics and Gender	Views toward sex differences in mathematical ability and the need to know mathematics for career purposes.	4
7.	Mathematics as a Process (b)	View of the nature of mathematics as a discipline...as a set of rules or a field where creativity, speculation, conjecture, and heuristics are important; as a field with fixed or changing content.	15

(a) The first three items of this scale, and an additional item, were presented to teachers. These items were not included in the present study due to the questionable selection and comprehensiveness of these four items.

(b) Presented to both students and teachers.

Table 2

Stratification and Selection Procedures
for the Sample Classes

	Grade 12 B.C.	Grades 12 & 13 Ont.	Pre-calc. & Calc. U.S.
Population Definition	Grade 12 students in public schools enrolled in Algebra 12.	1. Grade 12 students in public and private schools enrolled in Grade 12 Mathematics 2. Grade 13 students in public and private schools enrolled in at least two of Relations, Calculus, and Algebra	Students in public and private schools enrolled in 4th year mathematics courses with prerequisites of three years of secondary level mathematics (Algebra and Geometry).
	Students in private schools were excluded (approx. 3% of Grade 12 population).	Students in special schools for foreigners and schools with no fixed timetable were excluded.	
Stratification	Geographic regions regularly used by Ministry of Education, and school size.	Geographic region, size of community, public/private, English/French ration of Grade 13 and Grade 12.	Public/Private regional standard metropolitan statistical area, status code.
Selection of	A) Proportional allocation of classes to strata		Proportional allocation of classes to strata.
	b) Allocation of sample to schools categorized by size.		Allocation of sample to school districts categorized by size.
	c) Allocated number of schools selected proportional to size.	Five schools selected proportional to number of Grade 13s.	Two schools proportional to size.

d) One or, in a few cases, two or three randomly selected from samples schools.	One Grade 12 class plus one class from each of Relations, Algebra, randomly selected from sampled schools.	Two classes randomly selected from sampled schools. (Private class sample nationally in two stages: schools selected proportional to size and two classes randomly selected from sampled schools).
e) All students in selected classes.	All students in selected classes.	All students in selected classes.

Source: Garden, R. A. (1985).

worded items was reversed. The result was a consistent, matched student within class-teacher file. The final sample sizes for each country are summarized in Table 3.

Internal consistency. Before proceeding to the statistical analyses conducted to investigate the research hypotheses, item analysis (Nelson, 1974) were performed for each scale and sample. The results are reported in Table 4.

Examination of these data reveals that the properties of the items and scales are quite comparable among the five student samples and for the teacher sample on the one scale. Furthermore, these results are similar to the results from the international trails (see Kifer, 1979), and analysis only at the item level is warranted for Calculators and Computers and Mathematics as a Process. Consequently, given the desirability of a uniform approach to analyses across scales, "item" was included as a factor along with country.

Unit of Analysis. Related to the previous decision was the question of which unit of analysis--student's class--should be used to examine the country and item factors. Much controversy surrounds this issue (see, for example, Hopkins, 1982).

In line with the approach taken by Kifer (Travers, to appear) item scores were aggregated to the class level. It was felt that, because testing took place at the end of the school year, the assumption of independence among students within class was difficult to justify. On the other hand, the assumption of independence between classes and, particularly between teachers within schools, was held to be tenable. Initial examination of between class and between teacher differences in schools where more than one class-teacher was assessed revealed distinct differences. As a consequence,

Table 3

Final Student, Class, and Teacher Sample Sizes

Teachers	Country	Number of Students	Number of Classes	Mean	s.d.	Class Size Range
Grade 12	B.C.	1943	95	20.5	6.8	3-35
Grade 12	Ont.	1236	55	22.5	5.6	9-37
Pre-Calculus	U.S.	3891	207	18.8	7.4	3-45
Grade 13	Ont.	3143	175	18.0	6.9	1-32
Calculus	U.S.	731	43	17.0	6.3	3-30

Table 4

Mean, Standard Deviations, Internal Consistencies, and Range of Item-Scale Correlations

	No of.			Item	Ite:	Internal Con. & Ranges Item-Scale
1. Mathematics in School Importance	15	12 B.C.	3.65	0.45	.79	.03-.58
		12 Ont.	3.59	0.52	.83	.02-.61
		Pre-C. U.S.	3.75	0.49	.82	.14-.59
		13 Ont.	3.77	0.51	.82	.10-.60
		Calc. U.S.	3.87	0.45	.80	.13-.60
Difficulty	15	12 B.C.	3.35	0.45	.80	.24-.54
		12 Ont.	3.33	0.48	.81	.32-.54
		Pre-C. U.S.	3.41	0.45	.79	.29-.51
		13 Ont.	3.44	0.50	.82	.28-.55
		Calc. U.S.	3.55	0.49	.81	.29-.54
Liking	15	12 B.C.	3.09	0.47	.78	.03-.50
		12 Ont.	3.02	0.51	.51	.12-.51
		Pre-C. U.S.	3.09	0.49	.78	.14-.47
		13 Ont.	3.14	0.51	.78	.08-.52
		Calc. U.S.	3.13	0.49	.77	.12-.51

2. Calculators and Computers	8	12 B.C.	3.53	0.50	.62	.00-.47
		12 Ont.	3.50	0.45	.48	-.11-.36
		Pre-C. U.S.	3.64	0.46	.53	-.02-.39
		13 Ont.	3.55	0.48	.54	.00-.42
		Calc. U.S.	3.65	0.44	.51	.04-.40
3. Home Support for Mathematics	9	12 B.C.	3.34	0.64	.76	.31-.53
		12 Ont.	3.42	0.63	.73	.31-.55
		Pre-C. U.S.	3.54	0.60	.73	.34-.56
		13 Ont.	3.33	0.62	.74	.26-.55
		Calc. U.S.	3.52	0.58	.72	.26-.58
4. Mathematics for Me	19	12 B.C.	3.57	.087	.87	.01-.72
		12 Ont.	3.45	0.64	.91	.27-.77
		Pre-C. U.S.	3.69	0.56	.89	.21-.72
		13 Ont.	3.62	0.59	.90	.27-.71
		Calc. U.S.	3.79	0.53	.88	.24-.74
5. Mathematics and Utility	8	12 B.C.	3.46	0.59	.78	.40-.56
		12 Ont.	3.63	0.63	.78	.43-.56
		Pre-C. U.S.	3.91	0.55	.74	.36-.52
		13 Ont.	3.71	0.59	.75	.39-.54
		Calc. U.S.	3.95	0.54	.74	.35-.53
6. Mathematics and Gender	4	12 B.C.	3.80	1.00	.87	.59-.81
		12 Ont.	3.90	0.85	.81	.49-.69
		Pre-C. U.S.	3.85	0.93	.83	.52-.74
		13 Ont.	3.83	0.90	.81	.52-.69
		Calc. U.S.	3.85	1.00	.87	.59-.79
7. Mathematics as a Process	15	12 B.C.	3.16	0.37	.58	.12-.35
		12 Ont.	3.19	0.35	.57	.04-.40
		Pre-C. U.S.	3.28	0.35	.57	.06-.37
		13 Ont.	3.29	0.37	.61	.10-.39
		Calc. U.S.	3.35	0.37	.63	.05-.44
		II. Teachers				
Mathematics as a Process	15	12 B.C.	3.64	0.33	.55	-.07-.40
		12 Ont.	3.65	0.33	.61	-.02-.48
		Pre-C. U.S.	3.63	0.41	.74	-.04-.50
		13 Ont.	3.66	0.36	.65	.08-.53
		Calc. U.S.	3.62	.036	.63	0.02-.43

(a) Hoyt (1941)

students were aggregated to the class level, yielding a class-by-item matrix for each scale and country.

The class median was used as the aggregated class-item score. Resistant to outliers, it was felt that the median would provide a more valid measure of class performance than the more typically used mean.

Statistical analysis. A 5 x k (country-by-item, k the number of items) median polish (Tukey, 1977; Welleman & Hoaglin, 1981) was used to analyze the unique and joint effects of country and item. Similar to analysis of variance, the median polish is based upon the additive model, but fits the model by finding row and column medians and by using iteration to obtain a final solution. Row effects indicate the extent to which countries responded more or less positively to the k items in a scale. Column effects correspond to item effects and indicate which items were responded to more or less favourably. Finally, cell entries contain the residuals. Countries or items which fail to follow a general pattern established by other countries or items will produce residuals. These represent unique patterns of response by students in a particular country to particular items.

The median polish was completed separately within each scale using the computer program Minitab (Ryan, Joiner, & Ryan, 1982) with two complete iterations.

As is the case for other exploratory data analysis techniques, the median polish does not have an accompanying statistical hypothesis-testing procedure to identify significant effects. Instead, Tukey (1977) recommends the use of judgment, taking into account the nature of the distribution of effects. Examination of the distributions in the present study revealed that many of the effects were either equal to zero or close to zero. Application of a rule of thumb based on hinges and multiples of the H-spread suggested by Tukey (1977, p. 383) led to inconsistent findings across item effects and country-by item effects. In some instances the largest item effect was within the upper and lower hinges for items, while country-by-item effects of smaller magnitude were outside these hinges of residuals. To try to clarify the situation, a 5 x k fixed effects analysis of variance was performed in which "item" was treated as a repeated measures factor. The effect sizes yielded by this analysis were very similar in magnitude to those produced by the median polish. And, as with Tukey's rule, inconsistent findings were observed: small, near zero effects were significant (p. 01, Greenhouse-Geisser and Hunyh-Feldt probabilities (Kirk, 1982, pp. 259-262) in the case of students, but not for teachers. The use of class medians rather than individual student scores accounts for this increase in power.

Therefore, favouring a uniform procedure, the following rule was used: effects less than 0.20 in absolute value were considered non-significant and equal to zero. This cut-off corresponded in most instances to a natural breakpoint in the distributions of median polish effects, reflected fairly the skewness of these distributions where it existed, and represented a difference of 10 percent, the minimum considered necessary for interpretation.

To help clarify the presentation and discussion of results, the scales have been divided into three groups. The first group deals with the mathematics curriculum and contains the Mathematics in School scales and the Calculators and Computers scale. The second group centres on personal perceptions of mathematics, and included the Home Support, Mathematics and Me, Mathematics and Utility, and Mathematics and Gender scales. Lastly, the third group concentrates on views of the discipline of mathematics and contains the Mathematics as a Process scale.

Student Perceptions of the Mathematics Curriculum

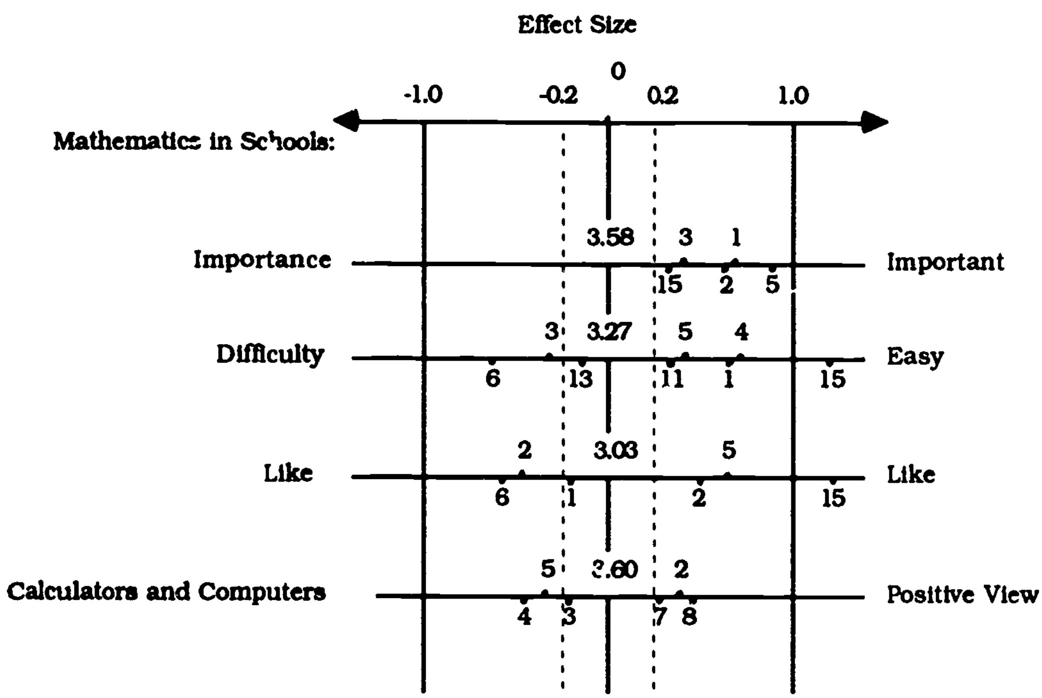
Results on the Calculators and Computers items indicated that, generally, Grade 12 Students in B.C., Grade 12 and 13 students in Ont., and Pre-Calculus and Calculus students in the U.S. were positive about the efficacy and benefits of calculators and computers (median item score = 3.60). They considered the 15 topics and activities in the Mathematics in School scale to be somewhat important (m = 3.58). They were, however, undecided about the difficulty of many of the topics and activities (m = 3.03)-

These overall findings held up across the five countries. As might have been predicted from consideration of the means listed in Table 4, the country effects produced by median polishing the corresponding country-by-item matrices were all less than 0.20 in absolute value.

Examination of the country-by-item effects revealed essentially the same finding. Of the 265 residuals, only 33 exceeded the significance criterion, and of these, only 16 could be reasonably explained. Not unexpectedly, Grade 13 students in Ont., and to a greater extent, Calculus students in the U.S. indicated that differentiating and integrating functions (Items 10, 13, Mathematics in School) were activities which were more important, relatively easy, and best liked. The U.S. Calculus students also considered drawing graphs of functions (11) and finding a limit of a function (12), two related activities, to be more important. They also indicated proving theorems (6) was somewhat less important and they tended to dislike this topic. This set of findings is attributable to the differences in curriculum between the countries. Grades 12 students in B.Cand Ont. and Pre-calculus students in the U.S. are typically not exposed to calculus. Students in Grade 13, Ont. generally are exposed to a variety of advanced topics (trigonometry, geometry, advanced algebra, calculus) while the calculus students concentrate for the most part on calculus. It seems likely that the negative feelings expressed by U.S. calculus students toward proving theorems can be attributed to their recent experience with proof in calculus.

The remaining significant residuals either failed to form meaningful patterns or were not easily explained. For example, it is not clear why, in comparison to the other students, Grade 12 students in Ont. found memorizing rules and formulas (2) more important, relatively less difficult, and more to their liking. Nor is it clear why Grade 12 students in B.C. particularly enjoyed investigating sequences and series (9), why U.S. Calculus students found determining the probability of an outcome (13) difficult, or why Grade 13 students in Ont. considered getting information from statistical tables (4) less important. Thus, except for the country-by-item effects explained by exposure to calculus, students from B.C., Ont., and the U.S. had similar perceptions of the topics and activities presented.

Item differences. In contrast to the absence of country differences, there were several significant item effects. These items are shown in Figure 1. The median item scores listed at the zero effect point provide a reminder of the overall position of each item set.



Polarity for negatively worded items has been reversed.

Figure 1. Items Related to the Mathematics Curriculum

1. For all three groups, the findings presented here are based upon the detailed results available from the authors.

Two of the five items considered most important were checking an answer by going back over it (1) and memorizing rules and formulas (2). These two activities were also two of the three least liked. The importance assigned to these tedious, relatively disliked activities may be explained by the practical orientation of the students. As will be seen later, student responses to Mathematics and Utility items suggested that the students do not study mathematics because they like it or find it intrinsically interesting, but rather because of its practicality.

The greater importance attached to solving equations (5) and solving word problems (3) is also consistent with this practical view of mathematics. The students did, however, react differently in their assessment of the difficulty of these activities; solving equations is easy, while solving word problems is more difficult. This difference may reflect the greater complexity of solving word problems; a problem must first be understood and correctly translated into an equation which then must be solved. This notion that more complex activities are perceived to be more difficult than simple straightforward activities can also be seen in the ratings of the difficulties of the following four activities: "proving theorems" (6) and "integrating functions" (13)--more difficult; "getting information from statistical tables" (4) and "drawing graphs of functions" (11)--less difficult (easy).

The fifth most important activity, "using a hand-held calculator" (15), was also the easiest and best liked activity of the 15 considered. When asked to react to specific issues related to calculators and computers, the students were equally, but judiciously, enthusiastic. They disagreed that calculators eliminated the need to learn to compute (2, Calculators and computers), and they felt that calculators were not particularly useful in learning different mathematical topics (3). The use of calculators did not ameliorate their dislike for solving word problems (4) (suggesting that the interpretation and translation of word problems is what students most dislike). The students agreed that computers were beneficial (5,8), and endorsed the suggestion that "everybody should learn something about computers" (7). These findings are congruent with the prominent role played by calculators and computers in a modern technological society, reflect the practical orientation of the students, and are indicative of the strong emphasis being given to learning about and how to use calculators and, especially, microcomputers in today's schools.

Personal Perceptions of Mathematics

In general, Grade 12 students in B.C., Grade 12 and 13 students in Ont., and Pre-calculus and Calculus students in the U.S. shared the same perceptions of their parents' ability in and support for mathematics, enjoyed to the same extent and felt equally competent studying mathematics, felt the same way about the importance of mathematics in preparing for an occupation and the usefulness of mathematics in everyday life, and held common views about the mathematical capability of boys and girls.

Two country effects exceed the minimum criterion for significance. Both Pre-calculus and Calculus students in the U.S. were stronger in their view that mathematics is important in preparing for a job and in solving everyday problems. Mathematics and the sciences enjoy a relatively high profile in the U.S. Considered a world leader in scientific advances and industrial development, mathematics and science are continually stressed. National ills of the country are often traced to the failure of the schools, and frequently to the failure of the schools to provide an adequate education and training in mathematics and science. The magnitude of involvement in like activities in Canada and the competitiveness of Canadians appears not to be as great.

Thirteen of the 200 country-by-item effects were significant. Again, not all appear to be meaningful. Of the 13, only six could be reasonably explained. U.S. Calculus students perceived their mothers as enjoying mathematics less and as less capable of assisting them with their homework (2, 4, Home Support). Given that these students were studying calculus, and that fewer women than men in the past studied mathematics beyond senior high school, and therefore, calculus, these findings are not surprising.

The U.S. Calculus students were more confident of their own ability to do mathematics (6, 11, Mathematics and Me), and to become good mathematicians (12). Presumably among the most able students in school, they strongly looked forward to taking more mathematics (4).

Item differences. The median polish yielded several significant item effects, particularly for items in the Home Support and Mathematics and Me scales. The significant items are shown in Figure 2.

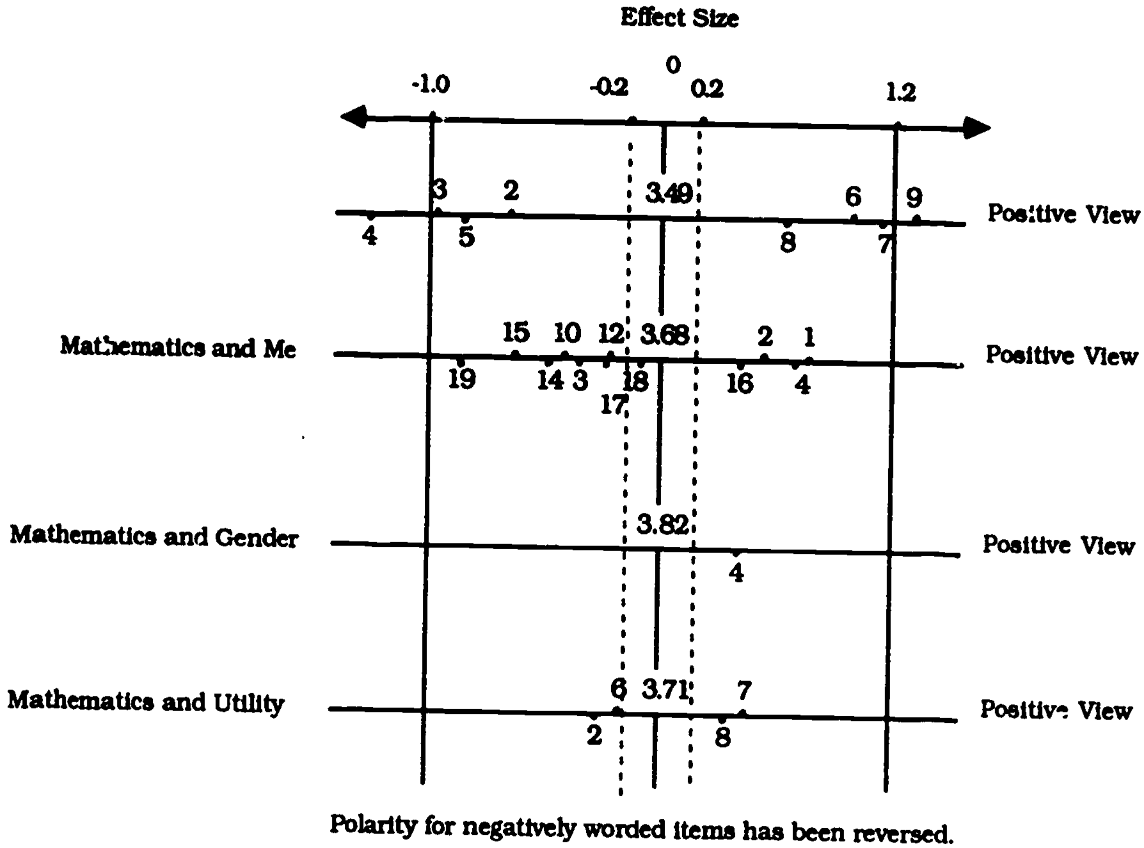


Figure 2. Items Related to Personal Perceptions of Mathematics

As shown in the Home Support scale, the students felt that their parents considered mathematics to be an important subject for them (the students) to study (6, 7), and that their parents encouraged them in their mathematical studies (8, 9). They did, however, feel that their parents usually were not very interested in helping them with mathematics (5). They questioned the ability of both their fathers (3) and, especially, their mothers (4) to do their homework, and indicated that their mothers tended not to enjoy mathematics (2). If these latter perceptions are accurate, then the observation that their parents, while supportive, are disinterested in assisting them with their work is understandable. It seems apparent that the students believed that the mathematics they were studying was beyond that studied by their parents. Still this did not appear to diminish the positive disposition of the parents toward their children's study of mathematics or their desire for their children to do well (9).

The students' perceptions about themselves were less clear. They wanted to do well in mathematics (1, Mathematics and Me). In general, they felt competent, but, with the exception of U.S. Calculus students, the students were uncertain that they could ever become good mathematicians (12). They were also undecided as to whether they were looking forward to taking more mathematics (3). Furthermore, they were unsure about spending a lot of their own time doing mathematics (10) and working for a long time to understand new ideas (14). Confronted with a problem they could not solve, they reported that they felt "lost in a maze" from which they could not find their way out (19). Yet, when they solved a problem, they felt good (4). Though mathematics did not make them "happy" (15), nor was it "fun" (18), the students did not fear taking mathematics (16).

Taken as a whole, these findings are not too surprising. They are consistent with what would be expected from students who felt they "had" to take mathematics. The high retention rates and graduation requirements of Canadian and American schools result in more students than just the academically able taking senior level mathematics. For the majority, mathematics may be more a means to an end, and not an end in itself.

This conjecture is supported by the effects observed for the Mathematics and Utility items. There was general agreement that mathematics was needed in everyday life (4, 5, 7). The students further agreed that knowledge of mathematics is necessary for most occupations (8), although they were not as sure that most mathematics had practical use on the job (6), or that most people actually used mathematics in their work (2). It appears the students believed that, in order to get a job, it was necessary to study mathematics, but what was actually covered was not always relevant to what was needed. Support for this interpretation can be seen in the differential importance assigned to some of the topics and activities of the mathematics curriculum. Moreover, this helps explain some of the indecision noted in the students' self-perception.

The students displayed a high degree of support for the equality of boys and girls. They agreed that a woman needs a career as much as a man (4, Mathematics and Gender), and that there were no differences between boys and girls in their ability in and need for mathematics.

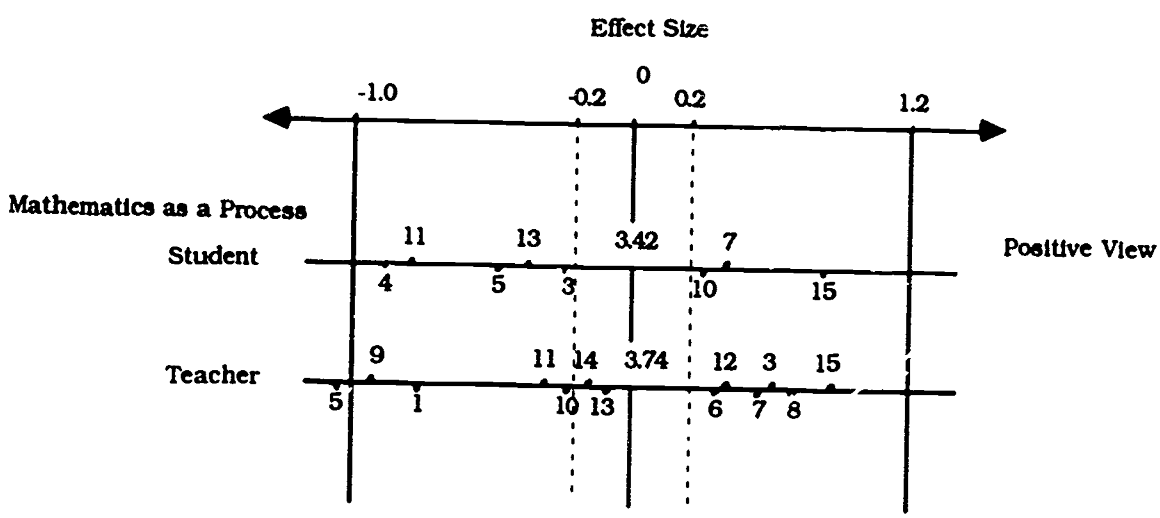
Student and Teacher Perceptions of Mathematics as a Process

The students were, in general, uncertain about the nature of mathematics as a field of study (median item score 3.42). Their teachers, while not always consistent, were generally more decided (m = 3.74). No country effects were found, and, except for a consistent country-by-item effect which revealed teachers in B.C. were less rule oriented (5, 9, 10, 11), no meaningful residuals were observed. As before, there were item differences for both students and their teachers.

The teachers were somewhat inconsistent in their view of mathematics as a changing field. They agreed that there had been recent discoveries in mathematics (12), but were undecided about changes in the near future (1). While more consistent, the students were essentially undecided about whether or not mathematics is a changing field.

Teachers generally agreed that mathematics provided the opportunity for originality (3, 8). The students were less sure. The students tended to disagree that learning mathematics involved mostly memorizing; the teachers clearly disagreed (8).

Both teachers and students agreed that "mathematics helps one thing logically" (15). When asked if "mathematics helps one think according to strict rules" the teachers agreed, while the students were undecided (5). The students were clearly undecided about whether or not mathematics was a set of rules; their teachers tended to disagree (13). The students though, were more rule oriented in their solution of mathematics problems (9, 11). Somewhat contradictory to these rules, students tended to agree that trial and error can often be used to solve a problem, while their teachers were less decided (10).



Polarity for negatively worded items has been reversed.

Figure 3. Items Related to Mathematics as a Process

Taken together, these results suggest that, in general the teachers were more process oriented than their students. This finding is in keeping with the suggestion that the greater the experience, the greater the process orientation. But the lack of a process-oriented view of mathematics by the students is somewhat puzzling. As senior level students, they ostensibly have had a fair amount of experience in mathematics. This leads to questions about the type of experience they have and the way in which mathematics is taught. It may be that the students, with their practical orientations, focussed on answering a problem correctly by the "right" rule, and that they cared little about how rules operate or from where they came. All that was needed was to know the right one and how to apply it. Teachers, with more mathematics education and experience, appear to be more insightful about the derivation and use of rules. It seems, though, that their teaching may be less process oriented, with stress placed on a "right rule--right answer" approach.

Summary

Overall, the findings presented and discussed support the similarity hypothesis suggested in the introduction, and reflect a practical view of mathematics. Grade 12 students in B.C., Grade 12 and 13 students in Ont. and Pre-calculus and Calculus students in the U.S. indicated practicality, and non-intrinsic worth, as the reason for studying mathematics. For the majority, mathematics appeared to be a means to an end, and not as end in itself.

Consistent with this view, the students considered the 15 curriculum topics and activities presented to be important, but they were unsure of their difficulty and less likely to like them. The students indicated that, although they would take more mathematics, they were unwilling to commit much of their "own" time in studying mathematics, and felt uncomfortable with new problems. Instead, they saw mathematics not so much as a field involving speculation and conjecture, but as a field in which problems were solved by a "learned, right" rule.

These results are disappointing but understandable. It is to be hoped that students in senior mathematics class would have a more process oriented, somewhat less utilitarian view of mathematics. This is not to say that practicality does not have a place; rather it is a question of balance. Why this balance was not more evident is attributable, at least in part, to the prevailing opinions held by many that mathematics is a service course, and to the way in which it is likely taught. The mathematics curriculum, as presently structured, favours a more linear, systematic approach, with little room for considering the development of mathematics as a field of study.

If these perceptions are indeed accurate, then students will need not only a "how to do it" acquaintance with mathematics, but also greater understanding of its place in a rapidly changing technological society, both in terms of its impact and its potential. Helping students to explore the nature of mathematics, as well as how to do it, is an important aspect of the development of a mathematically literate society.

References

Andrews, J. H. M., & Rogers, W. T. (1982). Canadian Research in Education: A sSate of the Art Review. Ottawa, Ont.: Minister of Supply and Service.

Garden, R. A. (1985). Second IEA Mathematics Sampling Report. Champaign, IL: University of Illinois (unpublished report).

Hopkins, K. D. (1982). The unit of analysis: Group means versus individual observations. American Educational Research Journal.

Husen, T. (Ed.). (1967). International Study of Achievement in Mathematics. Stockholm: Almquist & Wiskell.

Kifer, E. (1979). A report of the attitude trials for the 2nd IEA mathematics study. Lexington, Ky: University of Kentucky.

Kirk, R. E. (1982). Experimental Design: Procedures for the Behavioral Sciences (2nd ed.). Belmont, Calif.: Wadsworth.

Nelson, L. R. (1974). Guide to LERTAP Use and Interpretation. Dunedin, N.Z.: University of Otago Press.

Peaker, G. F. (1975). An Empirical Study of Education in Twenty-one Countries: A Technical Report. Stockholm: Almquist & Wiksell.

Ryan, T. A., Joiner, B. L., & Ryan, B. F. (1982). Minitab Reference Manual. University Park, PA: Statistics Department, Pennsylvania State University.

Travers, K. and Westbury, I. (Eds). (1989).. The IEA Study of Mathematics I: Analysis of Mathematics Curricula. Oxford: Pergamon Press.

Tukey, J. W. (1977). Exploratory Data Analysis. Reading, Mass.: Addison-Wesley.

Velleman, P. F., & Hoaglin, D. C. (1981). Applications, Basics, and Computing of Exploratory Data Analysis. Belmont. Ca.: Wadsworth.

Walker, D. A. (1976). The IEA Six Subject Survey: An Empirical sSudy of Education in Twenty-one Countries. Stockholm: Almquist & Wiksell.

Appendix

Items and Scales

Mathematics in School

1. Checking an answer to a problem by going back over it.

(a)	very important	important	undecided	not important	not at all important
(b)	very easy	easy	undecided	hard	very hard
(c)	a lot	like	undecided	dislike	dislike a lot

2. Memorizing rules and formulas. (response categories for all remaining items in this group are as shown for the first item)
3. Solving word problems.
4. Getting information from statistical tables.
5. Solving equations.
6. Proving theorems.
7. Using vectors.
8. Working with complex numbers.
9. Investigating sequences and series.
10. Differentiating functions.
11. Drawing graphs of functions.
12. Finding a limit of a function.
13. Integrating functions.
14. Determining the probability of an outcome.
15. Using a hand-held calculator.

Calculators and Computers

(Items marked * in this and the remaining scales are negatively worded)

*1. It is less fun to learn mathematical ideas if you use a hand-held calculator.

strongly disagree	disagree	undecided	agree	strongly agree

*2. If you use a hand-held calculator you do not have to learn how to compute. (Response categories for the remaining items in this and other scales are as shown for the first item.)

3. Using a hand-held calculator can help you learn many different mathematical topics.

4. Solving word problems is more fun if you use a hand-held calculator.
*5. Computers solve problems better than people do.
*6. Using computers makes learning mathematics more mechanical and boring.
7. Everybody should learn something about computers.
8. Computers do lots of good things for people.

Home Support for Mathematics

1. My father seems to enjoy doing mathematics.
2. My mother seems to enjoy doing mathematics.
3. My father would usually be able to do my mathematics homework problems if I asked him for help.
4. My mother would usually be able to do my mathematics homework problems if I asked her to help.
5. My parents are usually very interested in helping me with mathematics.
6. My mother thinks that learning mathematics is very important for me.
7. My father thinks that learning mathematics is very important for me.
8. My parents encourage me to learn as much mathematics as possible.
9. My parents want me to do very well in mathematics class.

Mathematics and Me

1. I really want to do well in mathematics.
2. My parents really want me to do well in mathematics.
3. I am looking forward to taking more mathematics.
4. I feel good when I solve a mathematics problem by myself.
5. I usually understand what we are talking about in mathematics class.
*6. I am not so good at mathematics.
7. I like to help others with mathematics problems.
*8. If I had my choice I would not learn any more mathematics.

9. I feel challenged when I am given a difficult mathematics problem.
*10. I refuse to spend a lot of my own time doing mathematics.
*11. Mathematics is harder for me than for most persons.
*12. I could never be a good mathematician.
*13. No matter how hard I try I still do not do well in mathematics.
14. I will work a long time in order to understand a new idea in mathematics.
15. Working with numbers makes me happy.
*16. It scares me to have to take mathematics.
17. I usually feel calm when doing mathematics problems.
18. I think mathematics is fun.
*19. When I cannot figure out a problem, I feel as though I am lost in a maze and cannot find my way out.

Mathematics and Utility

1. It is important to know mathematics in order to get a good job.
*2. Most people do not use mathematics in their job.
3. I would like to work at a job that lets me use mathematics.
4. Mathematics is useful in solving everyday problems.
*5. I can get along well in everyday life without using mathematics.
6. Most of mathematics has practical use of the job.
*7. Mathematics is not needed in everyday living.
*8. A knowledge of mathematics is not necessary in most occupations

Mathematics and Gender

*1. Men make better scientists and engineers than women.
*2. Boys have more natural ability in mathematics than girls.
*3. Boys have to know more mathematics than girls.
4. A woman needs a career just as much as a man does.

Mathematics as a Process

1. Mathematics will change rapidly in the near future.
2. Mathematics is a good field for creative people.
*3. There is little place for originality in solving mathematics problems.
4. New discoveries in mathematics are constantly being made.

*5. Mathematics helps one to think according to strict rules.
6. Estimating is an important mathematics skill.
7. There are many different ways to solve most mathematics problems.
8. Learning mathematics involves mostly memorizing.
9. In mathematics, problems can be solved without using rules.
10. Trial and error can often be used to solve a mathematics problem.
*11. There is always a rule to follow in solving a mathematics problem.
*12. There have not been any new discoveries in mathematics for a long time.
13. Mathematics is a set of rules.
14. A mathematics problem can always be solved in different ways.
15. Mathematics helps one to think logically.

The research reported in this paper was supported by a grant from the Social Science and Humanities Research Council of Canada (No. 410-83-0702). We would like to thank Robert Prosser for his able assistance in carrying out the data analysis.

STATE CONTROL OF THE CURRICULUM: POLITICAL INCORPORATION AND IMPLEMENTATION OF THE CURRICULUM

David L. Stevenson
Office of Research
U.S. Department of Education

David P. Baker
Department of Sociology
Catholic University of America

The emergence of the modern nation-state and the emergence of mass education are closely intertwined. The development of modern nation-states relied, in part, upon several functions of formal schooling, such as the creation of citizens, the establishment of a legitimated system of economic and political allocation and the socialization of a labor force for a national economy. At the same time, agencies of the state provided resources for funding and chartering of educational expansion and, thereby, influenced the organization and content of educational activities.

In this paper, we investigate an aspect of the relationship between the state and schooling, the state's control of the curriculum. We examine whether national state regulation of the curriculum is related to curriculum implementation in the classroom.

The linkages of macrosociological characteristics, such as state control, to microsociological characteristics, such as implementation of the curriculum, are seldom studied because of extensive data requirements. To examine such an issue, we have created a large comparative data set of 15 educational systems with information on the political incorporation of education as well as implementation of curriculum in the classroom.

Political Incorporation of Curriculum Control

In assessing the relationship between the state and education, Ramirez and Rubinson (1979) contend that world-wide growth in state authority and power increases the political incorporation of education. They suggest that the political incorporation of education can explain several recent trends in education.

---

*This is an early draft of this paper. The final version will be published in Sociology of Education and that version of the paper should be cited.

One major trend is the world-wide expansion of formal schooling as measured by enrollment rates. Ramirez and Rubinson find that the state's authority and power are clearly related to growth in enrollments in all public sectors (primary, secondary and tertiary) of schooling. They also argue that the political incorporation model generally fits results from other studies of the growth in enrollments (e.g., Boli, Ramirez & Meyer, 1985; Meyer & Hannah, 1979) and that there is a lack of empirical support for either human capital or status conflict accounts of educational expansion (Rubinson, 1986).

A second major trend is the growth in the number of educational systems with compulsory schooling laws. A recent study of the compulsory schooling laws in the 19th century indicates a relationship between political incorporation and the passage of compulsory schooling laws (Ramirez & Boli, 1987).

What has not been adequately studied is whether the political incorporation of education influences educational activities in the classroom. We examine this issue for one significant educational activity, implementation of the curriculum in the classroom. Ramirez and Rubinson (1979) discuss the proposed research question as a needed critical test of the political incorporation model of the relationship between the state and education.

Official Curriculum and the Implemented Curriculum

There is a renewed interest among sociologists in the study of state control and the content of the official curriculum. For example, an area of considerable interest is the changing content of national curriculum and the process by which a curricular subject is defined and institutionalized as a legitimate subject (e.g., Goodson, 1988; Goodson & Ball, 1984). These studies examine the historical development of the curricula, with an emphasis on how local politics shape the contents and definition of the official curricula (Apple, 1979). From an institutional perspective, others have analyzed the increasing homogeneity in the subject composition of official curricula of national educational systems from 1920 to 1985 (Benavot & Kamens, 1989; Benavot, Kamens, Wong & Cha, 1988).

Although these studies differ in their theoretical perspectives, they all focus on the <u>official curricula</u> of schooling. The official curricula is part of an elaborate classification system that defines the appropriate categories of instruction. Schools incorporate these categories into their organizational structure and activities (Meyer & Rowan, 1978). If the official curriculum requires the study of mathematics, schools create departments of mathematics, hire teachers of mathematics, and offer courses in mathematics.

Schools tightly control and monitor being in compliance with the subject categories of the official curriculum (Meyer, 1983). School officials are concerned that the curriculum 'fit' the state mandated curriculum. For example, they are concerned as to whether they offer the appropriate classes of algebra, world geography, and other subjects. By being in conformance with the categories of the official curriculum, schools maintain their legitimacy, gain access to resources and avoid sanctions, such as a loss of accreditation (Meyer & Rowan, 1978).

While schools tightly monitor their curricular offerings, there is variation in the degree to which there are organizational controls over the implementation of the official curriculum in the classroom.

Organizational Controls Over Instruction

Instruction is part of the technical activity of schools and one of the educational outputs of schooling. In systems that are more loosely coupled, such as in the U.S., educational organizations exercise weak bureaucratic controls over instruction. This is because in these systems the technical activities of schooling, instruction and learning, are buffered from inspection and assessment. Schools seldom attempt to assess these organizational outputs of schooling, in part, because of a lack of market pressures. The technical environments of schools in these types of systems do not provide significant constraints as neither the survival nor profitability of the school is determined by the quantity or quality of instruction. While schools keep elaborate records of certain types of educational outputs such as attendance, course enrollments, and number of graduates, they seek to avoid inspection of instruction. Thorough and frequent inspections of instruction may reveal inconsistencies and inefficiencies, thereby creating a challenge to existing organizational arrangements (Meyer, Scott & Deal, 198[illegible]). Teachers, in these types of systems, have a great deal of autonomy and discretion in the handling of instruction and learning. They often modify the official curriculum to meet their needs or those of their students, and, therefore, teachers teaching the same subject within a school may differ in the amount of material covered, the type of topics covered, the amount of time spent on instruction, and the use of curricular materials.

In other educational systems, the control of the curriculum and its implementation is greater. Classroom processes in these systems are less buffered from external influence. The technical environments of these schools are more clearly defined and influence larger segments of educational activities.

We argue that the degree to which an educational system is incorporated into the state will influence the degree to which the technical environment of schooling is controlled. The more incorporation with the state the more control and less autonomy at the classroom level. There are a host of mechanisms through which the state can control the implemented curriculum. These range from concrete forms of social control, such as state inspection, monitoring teacher training and formal assessment of student achievement, to more indirect forms of control, such as shaping the definitions of instruction and socialization of teachers. Although we do not measure these mediating mechanisms here, we can assess the presence or absence of their combined influence on the implemented curriculum.

State Control of the Curriculum and Implemented Curriculum

Educational systems vary in the degree of political incorporation of curricular subjects and their content. In some educational systems, control over curricular issues is highly centralized and managed at the national ministry of education. In other educational systems, curricular issues are dealt with at the provincial or local level. The degree of political incorporation of curricular matters affects the degree of environmental specification of instruction.

If state control over the curriculum is located at the national level, the environment is less complex and there will be greater specification of instruction. Through the ministry of education, or some administrative counterpart, there is an administrative mandate for what the curriculum should be. Such a mandate may be reflected in the curricular guidelines, the training of teachers, the content of curricular materials, and items on student achievement tests.

The national educational agency also may institute a set of bureaucratic controls to assure implementation of the curriculum. For example, state inspectors may occasionally visit classrooms to assess the content of instruction or academic achievement tests may be used to determine how students are allocated to classes and curricula. The effects of such bureaucratic controls on classroom instruction, however, are not well documented and may create little more than procedural compliance.

In educational systems with local political control of the curriculum, the environment of teaching is more complex and there is less specification of instruction. The administrative mandate as to what teachers should teach is weaker; there will be greater diversity in the textbooks available for use by the teachers; and there will be greater diversity in the types of training available for teachers. Since schools receive local funding and rely upon community support, they are likely to be more responsive to local constituencies.

This discussion about state control, technical environments and the implemented curriculum suggest two hypotheses about political incorporation.

Hypothesis 1: The greater the degree to which education is incorporated within the state the greater control over the implementation of curriculum in the classroom, which will be reflected in more uniformity across implementation by teachers within a system.

Hypothesis 2: Political incorporation simplifies the technical environments of schools, thus in highly incorporated systems local factors of classrooms will not influence the implemented curriculum.

Data and Methods

Testing these hypotheses requires detailed data about classroom instruction in educational systems that vary in the degree to which education is politically incorporated. The Second International Mathematics Study (SIMS) undertaken by the International Association for the Evaluation of Educational Achievement (IEA) provides this type of data. This large data set represents a powerful analytic resource for cross-national study of education. The countries in which SIMS collected data represent a diverse set of societies in terms of their size, geographic location and level of development. The use of a standard sampling procedure within each country yielded high quality samples of classrooms. Extensive efforts were undertaken to assure that comparable data collection procedures were used in each educational system.

The SIMS data were collected in 20 educational systems.[1] Of the 20 educational systems represented in the SIMS data set, 15 had full classroom process questionnaires.[2] In each educational system, a four step, stratified-random sample of 8th grade mathematics classrooms were drawn.[3] This yielded over 2200 classrooms. For ea class, detailed information was collected from the teacher about the amount and type of instruction in mathematics during the year. For 157 items in mathematics, each teacher was asked whether or not they had taught such an item during the year. Teachers in each educational system were asked the same information about the same 157 items in mathematics.

For each educational system, a board of educational experts designated which of the 157 items in mathematics were part of the national curriculum in mathematics for 8th grade.[4] How much this so-called national curriculum overlapped with official curriculum in various parts of each educational system was not evaluated by SIMS. At the very least, the measure of national curriculum, which we used here, represents the largest possible set of mathematics skills that an 8th grade teacher would cover on average in the course of the year.

Description of Measures

The political incorporation of education, as Ramirez and Rubinson (1979) define it, refers to the extent of national control over schooling. They suggest that a valid measure of political incorporation is the level of political control over education. The more that control occurs at the national level, the more schooling is politically

1 We analyze national educational systems, except for Canada, which collected data separately in British Columbia and Ontario. Because of some minor differences in data collection in these two provinces, we analyze them separately.

2 SIMS in Hong Kong, Scotland, French Belgium and Nigeria did not include questions about the implementation of curriculum. The Flemish Belgium sample did, and we will use it to represent Belgium. Swaziland was dropped from the analysis because only one-fifth of the teachers completed this part of the instrument.

3 See Garden (1987) for a detailed description of the SIMS study.

4 In each country this board was made up of representatives from the ministry of education, the teacher's union, teachers and school district level administrators. The panel was asked to assess which of the items from the item pool would mostly likely be part of the standard 8th grade mathematics curriculum in their country. The Japanese ministry decided that the items were too easy for the bulk of its 8th grade students so 7th grade classrooms were sampled.

incorporated with the state. As an indicator of this construct we have slightly modified a scale developed by Ramirez and Rubinson (1979). We used a seven point scale and ranked each country in terms of the political level that had the greatest control over the curriculum: 1) local control, 2) local and provincial control, 3) provincial control, 4) local, provincial and national control, 5) local and national control, 6) provincial and national control, and 7) national control. In coding each system on this scale, we consulted standard reference sources (*International Encyclopedia of Education*, 1985; *International Handbook of Educational Systems*, 1983) as well as an IEA publication with descriptions of the educational systems (Travers & Westbury, 1989). Three raters independently scored each educational system on the scale. The level of agreement among the three raters was above 98%.

From the SIMS data we constructed several indicators of different dimensions of the implemented curriculum. First, we took the number of items in the national curriculum (as determined by the panels of educational experts) as an indicator of the size of a system's official mathematics curriculum. Second, for each educational system, we calculated the percentage of the national curriculum that a teacher taught during the year and calculated a mean and standard deviation as indicators of the amount of curriculum covered in the system and the variation in the amount of curriculum covered. Third, we calculated the percentage of teachers in each educational system who taught each of the items in the national curriculum. As an indication of agreement among teachers' implementation of curriculum we counted the number of items that were taught by either 90% or more of the teacher or 10% or less of the teachers.

Finally, we have measures of local factors which might influence the implementation of the curriculum for each class such as the range in the mathematics abilities of students, the level of mastery of mathematics, the age and sex of the teacher, number of years the teacher has been teaching as well as teaching mathematics. We also have measures of the number of periods of mathematics per week and the average length of a mathematics period.

Analysis Plan

First, we correlated measures of various dimensions of the implementation of curriculum with the indicator of state control of the curriculum. Next, we used a model of teacher coverage of the national curriculum and estimated this model with each system's data. There are several advantages in doing this type of analysis which is a standard approach to analysis of student or classroom data and national factors (Heyneman & Loxley, 1982; 1983). Since our hypotheses are about relationships between institutional characteristics of systems, we required indicators of curriculum coverage at the system level and therefore we do not combine all classrooms into one sample. This approach allows our analysis to incorporate differences in the size and nature of the national mathematics curriculum in each system. It also allows us to handle some of the minor differences in questionnaires and procedures that are almost inevitable in a comparative study of this size and complexity.

Results

In the first column of Table 1 are measures of the size of the 8th grade mathematics curriculum in each educational system. While all of the educational systems in the sample had 8th grade mathematics, the size of their curriculum varied. The sample mean was 125.1 items (or 80% of the 157-item pool), with a standard deviation of over 16 items. The range in size was substantial. Three educational systems (New Zealand, Japan and Hungary) had a large curriculum that covered approximately 140 items (or over 90% of the 157-item pool). At the lower end, Belgium (Flemish) and Luxembourg had curricula that covered approximately 95 items (or only 60% of the 157-item pool).

The second column in Table 1 shows the mean number of items of the curriculum that were taught during 8th grade by each system's teachers. Here there is considerable variation with a standard deviation of 20 items and a range of over 70 items. Japanese teachers taught the most, with a mean of 117.2 items (or 75% of the 157-item pool), while Canadian (British Columbia) taught the least, with a mean of 42.7 items (or only 27% of the 157-item pool).

The third column in Table 1 is the mean number of items taught as a percentage of the total number of curricular items. In none of the educational systems studied did the "average teacher" cover the entire 8th grade curriculum. The sample mean is 65% with a standard deviation of over 15 percentage points. There is also a large range in coverage, with teachers in Belgium (Flemish) and Japan providing instruction for over 80% of their curricula and teachers in British Columbia and The Netherlands providing instruction for under 45% of their curriculum.

TABLE 1

Curriculum in Mathematics

Educational system	Size of national curriculum	# Items taught	Standard deviation of items taught	% National curriculum taught
U.S.	128	93.6	20.5	73.1
England	146	98.7	26.8	67.6
The Netherlands	127	55.2	16.2	43.5
Belgium (FL)	95	81.0	15.0	85.3
New Zealand	148	98.9	21.3	66.8
Canada (BC)	127	42.7	16.0	33.6
Canada (Ontario)	118	87.1	16.6	73.8
Finland	124	81.5	15.6	65.7
France	108	84.6	7.9	78.3
Hungary	142	65.9 (86.0)[a]	35.8 (26.3)[a]	46.4 (60.0)[a]
Israel	118	70.0 (62.0)[b]	22.5 (19.1)[b]	59.3 (52.5)[b]
Japan	146	117.2	10.0	80.3
Luxemburg	97	71.7	10.9	73.9
Sweden	122	60.1	13.9	49.3
Thailand	131	103.2	15.4	78.8

a Classrooms only in the Budapest area.
b Classrooms only in the Reformed system (7-9th grade).

Even though all of the educational systems had 8th grade mathematics as a curricular subject, the data in Table 1 indicate that there is variation among these educational systems in the content of their mathematics curriculum. Also, the amount of instruction varies considerably across educational systems. While the school curricula may have become institutionalized at the world level, our data suggest that there remains systemic variation in content and instruction.[5]

---

5 Our analyses of these data do not indicate a ranking of an educational system's overall efficiency in mathematics instruction. We interpret the ranking only as an indication of variation in the "size" of and "conformity" to the official curriculum.

Our first hypothesis predicts that teachers in educational systems with state control of the curriculum at the national level would be more uniform in their implementation of the curriculum in the classroom. The results displayed in Table 2 indicate that in educational systems in which there is state control of the curriculum at the national level, there is a modest tendency for more uniformity in the number of items that teachers teach. The correlation between an educational system's standard deviation in the mean number of items taught and state control is negative and significant, but only after we make a minor correction for the Hungarian and Israeli samples. There is a stronger association between the minimum number of items taught in a classroom in an educational system and our indicator of state control. Education systems with state control of the curriculum at the national level tend to display less variation in the amount of instruction and do not have teachers who teach little of the curriculum.

TABLE 2

Correlations Between Political Incorporation and Implemented Curriculum

Mean number of items in national curriculum taught	Standard deviation of mean number of items in national curriculum taught	Least number of national curriculum items covered	Number of national curriculum items taught by <10% or >90% of teachers	Number of national curriculum items taught by <10% or >90% of teachers	Percentage of national curriculum items taught by >90% of teachers
-.10 (-.7)a	-.27 (-.48**)a	.46** (.58**)a	.39** (.59*)	.47** (.49**)	.45** (.45**)

** p< .05

a Coefficients in parentheses calculated with partial Israel and Hungary samples.

We also suggest that teachers in educational systems with state control of the curriculum at the national level would be more likely to teach the same material. To examine this issue, we constructed three indicators of the similarity among teachers in their classroom instruction and correlated these indicators with our measure of state control of curriculum. The first two measures are the number of items that 10% or less, or 90% or more, of the teachers in an educational system taught. These two measures indicate the extent of agreement in instruction among teachers. The first measure indicates the extent of agreement in coverage during the 8th grade year and the second measure indicates the extent of agreement during both the 7th and 8th grade years. Both of these agreement measures are moderately correlated with the level of state control of curriculum.

We constructed a third indicator of agreement that takes into account the variation in the size of the mathematics curriculum. We divided the number of items that at least 90% of the teachers taught in 7th or 8th grade by the number of items in the curriculum. The correlation between this measure and state control is similar in strength to the item counts. For each indicator, the analyses suggests that teachers were more likely to teach the same material if they taught in educational systems with national state control of the curriculum.

Our second hypothesis predicts that local factors will influence classroom instruction in educational systems with state control of curriculum at the local c provincial level To examine this issue we regressed the mean percentage of the national curriculum covered in 8th grade on indicators of local factors. The same equation was estimated for each sample of teachers and they are reported in Table :

If our description of the effects of state control are correct, we should find that the regression equations for educational systems with state control at the local or provincial level are significant. All of the educational systems with local curricular control had significant equations, while only two with national level control (Finland and Sweden) had significant equations. The correlation between the measure of state contrc of the curriculum and the squared multiple correlation coefficients resulting from the equations is -.67 ($p = .003$).

Among those educational systems with local state control of the curriculum, local factors account for from a low of 9% in the variation of instruction in The Netherlands to a high of 24% in England and Wales. In these educational systems a range of local factors predicted instruction. Teachers in these systems seem to be particularly sensitive to student resources within classrooms, both in terms of the average level of mathematical mastery of the class and the diversity of ability within the class. Following these factors, the amount of the instruction depends on time resources, both in terms of the number of mathematics sessions and the length of these sessions.

TABLE 3

OLS Regression of Local Factors on Implemented Curriculum

Educational system decentralized above the line	N	R^2	F	Intercept	Significant Local Factors							
					Student resources		Teacher resources				Time resources	
					Range of class	Mastery of class	Age	Sex	Experience teaching mathematics	Experience teaching	Periods per week	Average length of period
U.S.A.	253	.10	3.1	.61[a]		.0014				.0084		
England & Wales	204	.24	6.7	.17	-.041	.002	-.006					.0051
The Netherlands	206	.09	2.5	-.03	.031						.037	
Belgium	120	.16	2.3	.47							.065	
New Zealand	151	.19	3.8	.27		.0014			.0075	.0074		.0047
Canada (BC)	73	NS										
Canada (Ontario)	126	NS										
Finland	176	.12	2.7	.45		.0008						
France	286	NS										
Hungary	56	NS										
Israel	85	NS										
Japan	193	NS										
Luxemburg	79	NS										
Sweden	172	.09	2.8	.28		.0014						
Thailand	80	NS										

a The regression coefficients are unstandardized and significant at least with $p \geq .05$.

In Sweden and Finland, the two educational systems with national state control of the curriculum and significant equations, the overall level of mathematical mastery is the only significant variable. In both countries, there are numerous ability tracks in the 8th grade and classes in these tracks, by central administrative definition, should receive different amounts of mathematical instruction.[6]

We can explore two possible statistical artifacts within these results. One is the lack of significant regression equations for the educational systems with state control at the national level could result from a lack of variation in local factors. Educational systems with national control of the curriculum could also be the kind of educational systems that equalize between-classroom factors. The between-classroom local factors could be so similar that the non-significant equations result from a lack of between-classroom variation in local factors. To examine this possibility, we correlated the standard deviations of each of the eight indicators of local factors with the measure of state control of the curriculum. All of these correlations are small and not significant, except one. The exception is that educational systems with national state control of the curriculum tend to diminish the between-classroom variation in the number of mathematical instruction sessions per week (-.74, p=.001). In general, the between-classroom variation in local factors does not vary by the level of state control of the curriculum.

A second consideration is whether there is a spurious correlation between the level of state control of the curriculum and our various indicators of instruction. We have examined bi-variate associations at the system-level and there could be other system-level factors which might either mediate or negate the correlations that we report.

We examine four such factors -- two indicators of the economic development of the country (1980 Gross National Product and Gross Domestic Product), and two indicators of the size of the educational system (the population in 1980 and the gross primary enrollment ratio in 1980). We took the natural logarithm of each of these indicators and calculated a partial correlation between the measure of state control and the various indicators of instruction controlling for each of these factors.

[6] At this educational level, Finland has three ability tracks by classroom in mathematics (the short course, the long course, and the heterogeneous course) and Sweden has two ability groups by classroom (general and advanced).

Neither the indicators of the level of development or size correlated with the level of state control of the curriculum. Though there is a slight tendency for the larger educational systems to have local state control of the curriculum, the correlation is not statistically significant.[7] In each of the partial correlations the pattern of correlations between state control and instruction did not change after controlling for these other variables.

Discussion

Educational systems are linked to the state in a variety of ways and this is reflected in the degree of political regulation of education. For some educational systems there is national state regulation of educational activities through a ministry of education. While for other educational systems, educational activities may be unregulated or regulated at the local or provincial level. Our results indicate that this variation in state regulation of the curriculum is related to the implementation of the curriculum in the classroom. In educational systems with strong national control of curricular issues, we found that teachers were likely to teach the same material in the classroom. If there was local political control of curricular issues, the amount that teachers taught was determined by local factors.

These findings support the value of the political incorporation model as a general framework for examining the relationship of the state and education. The state can influence far more than the supply of educational opportunities and the chartering of schools. As we have shown, qualities of the state also can influence the core technical activities of schooling, classroom instruction. This lends additional credibility to studies of the state's role in forming the official curriculum. The influences of the state run from the creation of the official curriculum to its implementation in the classroom.

The state's control of curriculum and its implementation may increase worldwide. National political incorporation is fueled by internal processes of the state as

[7]The correlations between level of economic development and curricular coverage, and the size of the educational system and curricular coverage were generally not statistically significant. There are two implications of these findings. First, the lack of associations raises questions about hypotheses which suggest that curriculum coverage may be sensitive to economic and technical development (for example see Benavot & Kamens, 1989). And secondly, the lack of associations suggests that some structural characteristics of educational systems may not influence curricular implementation.

well as external forces. For example, consider the recent national and international discussions of the relative effectiveness of nations' educational systems (Lapointe, Mead & Phillips, 1989; McKnight, 1987). This debate illustrates the trend to consider student achievement as a national resource that should be, therefore, officially monitored by the state. These concerns encourage greater national political incorporation. As the 19th century state was concerned with expanding school enrollments and attendance, the late 20th century state is concerned with student achievement and teacher effectiveness.

Our findings have implications for several other lines of research. Observations about the weakness of organizational controls on classroom activities have perhaps overlooked the variation in political incorporation and its influence on the technical environment of schooling. Our results indicate that the degree of national incorporation of education is clearly related to classroom level activities; and institutional perspectives need to consider these findings.

These findings also have implications for the study of the relationship between schools and their environments. Results from other studies indicate that organizational characteristics of schools are related to characteristics of their environment. For example, the degree of administrative complexity in American public schools (Rowan, 1982) and public school districts (Meyer, Scott & Strang, 1987) is related to the degree of fragmentation of the environment and the formal structuring of environmental factors. This line of research has not, however, examined the relationship between the environment and educational outcomes. Our findings indicate that the complexity of the environment, as measured by the degree of national state regulation of the curriculum, is related to a significant educational outcome, classroom instruction.

We have examined only one curricular subject and additional research needs to be done on other areas of the curriculum. We predict that the relationship between political incorporation and implementation of the curriculum should be stronger for other subjects, such as civics and social studies. The content of these subjects is of greater interest to the state since they help to shape public definitions of citizenship and civic culture.

BIBLIOGRAPHY

Apple, M. (1979). *Ideology and Curriculum*. Boston: Routledge & Kegan Paul.

Benavot, A. & Kamens, D. (1989). "The Curricular Content of Primary Education in Developing Countries." Paper presented at the American Sociological Association Meetings.

Benavot, A., Kamens, D., Wong, S., & Cha, Y. (1988). "World Culture and the Curricular Content of National Education Systems, 1920-1980." Paper presented at the American Sociological Association Meetings.

Boli, J., Ramirez, F.O., & Meyer, J. (1985). "Explaining the Origins and Expansion of Mass Education." Comparative Education Review, 29, 145-170.

Garden, R. (1987). *Second IEA Mathematics Study: Sampling Report.* Washington, DC: U.S. Department of Education.

Goodson, I.F. (1988). *The Making of Curriculum: Collected Essays.* New York: Falmer.

Goodson, I.F. & Ball, S.J. (1984). *Defining the Curriculum: Histories and Ethnographies.* New York: Falmer.

Heyneman, S.P. & Loxley, W.A. (1982). "Influences on Academic Achievement Across High and Low Income Countries: A Re-analysis of IEA Data." Sociology of Education, 55, 13-21.

Heyneman, S.P. & Loxley, W.A. (1983). "The Effect of Primary-School Quality on Academic Achievement Across Twenty-nine High- and Low-Income Countries." American Journal of Sociology, 88, 1162-1193.

Lapointe, A., Mead, N. & Phillips, G. (1989). *A World of Differences: An International Assessment of Mathematics and Science.* Princeton: Educational Testing Service.

McKnight, C., et al., (1987). *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective.* Champaign, IL: Stipes.

Meyer, J. (1983). "Centralization of Funding and Control in Educational Governance", pp. 179-197. In J.W. Meyer & W.R. Scott (Eds.), *Organizational Environments: Ritual and Rationality.* Beverly Hills: Sage.

Meyer, J. & Scott, W. (Eds.). (1983). *Organizational Environments: Ritual and Rationality.* Beverly Hills: Sage.

Meyer, J., Scott, W. & Deal, T. (1983). "Institutional and Technical Sources of Organizational Structure: Explaining the Structure of Educational Organizations." pp. 450-67. In J.W. Meyer & W.R. Scott (Eds.), *Organizational Environments: Ritual and Rationality.* Beverly Hills: Sage.

Meyer, J.W., Scott, R. & Strang, D. (1987). "Centralization, Fragmentation, and School District Complexity." Administrative Science Quarterly, 32, 186-201.

Meyer, J.W. & Rowan, B. (1978). "The Structure of Educational Organizations," pp. 78-109. In M.W. Meyer and Associates (Eds.), *Environments and Organizations*. San Francisco: Jossey-Bass.

Meyer, J. & Hannan, M.T. (eds.) (1979). National Development and the World System. (Chicago: University of Chicago Press).

Meyer, J.W., Ramirez, F.O., Rubinson, R. & Boli-Bennett, J. (1977). "The World Educational Revolution, 1950-70." Sociology of Education, 50, 242-258.

Ramirez. F.O. & Boli, J. (1987). "The Political Construction of Mass Schooling: European Origins and Worldwide Institutionalization." Sociology of Education, 60, 2-17.

Ramirez, F.O. & Rubinson, R. (1979). "Creating Members: The Political Incorporation and Expansion of Public Education," pp. 72-82. In J.W. Meyer and M.T. Hannan (Eds.), *National Development and the World System*. Chicago: University of Chicago Press.

Rowan, B. (1982). "Organizational Structure and the Institutional Environment: The Case of Public Schools." Administrative Science Quarterly, 27, 259-279.

Scott, W.R. & Meyer, J.W. (1988). "Environmental Linkages and Organizational Complexity: Public and Private Schools," pp. 128-160. In T. James and H.M. Levin (Eds.), *Comparing Public and Private Schools, Vol. 1*. New York: Falmer.

Travers, K.J. & Westbury, I. (1989). *The IEA Study of Mathematics Volume I: Analysis of Mathematics Curricula*. New York: Pergamon.

APPENDICES

Appendix 1

Content Representation in College Algebra

Technical Appendix

Peter Lochiel Glidden

CONTENT REPRESENTATION IN COLLEGE ALGEBRA: TECHNICAL APPENDIX

Peter Lochiel Glidden
University of Illinois at Urbana-Champaign

February 1990

The Second International Mathematics Study College Algebra Classroom Process Data for Population B were examined: (a) to study reasons cited by teachers for teaching subtopics, (b) to study reasons cited for selecting particular content representations, and (c) to determine what relationships exist, if any, between teachers who use multiple content representations and their teaching decisions, professional opinions, backgrounds, classes, and schools. The major results[1] of this analysis are detailed in *Content Representation in College Algebra: Summary Report.*

Briefly, these results were: (a) External reasons and teacher familiarity frequently were cited as reasons for and against teaching particular topics in complex numbers and logarithms. Additionally, content reasons frequently are reported as reasons why a topic should be taught. Closely paralleling reasons for topic coverage, external reasons and teacher familiarity frequently were reported as reasons for and against using a particular concept representation and content reasons frequently were reported as reasons why a representation should be used. However, only for concept representation, **easy to understand** also was frequently reported as a reason for using a particular representation. For both subtopic coverage and concept representation, **easy to teach** and **enjoyed by students** were not often reported as reasons, either pro or con.

There were significant relationships between the use of multiple representations and teacher development and use of supplemental materials. There also was a relationship between multiple representation use and sources of information used to decide what to teach, how to teach, and what applications to present. Together these relationships suggest that teachers who use multiple representations (MRT) use more sources of information (self-developed materials, minimum competency statement, text, or syllabus) than do nonmultiple representation teachers (non-MRT).

Teachers who use multiple representations also allot more time for a topic and they are more likely to cover important formulas and theorems more deeply than nonmultiple representation teachers. There also was some evidence of a relationship between teacher experience/education and multiple representation use, but inferences can not be drawn at this time.

Purpose of Technical Appendix

The purpose of the Technical Appendix is to discuss briefly the isolated, statistically significant results of this study. Because this research was a first pass through the data

1 Major results are results that: (a) were supported by statistically significant relationships between at least two multiple representation indices and a variable and (b) had additional support from at least one other statistically significant relationship between at least one index and another closely related variable.

exploring relationships between multiple representation use and all the other classroom process data collected by SIMS, it is to be expected that statistically significant relationships will exist due to chance alone. Therefore, these results should not be interpreted as a picture of multiple representation use and teacher/ school characteristics, but rather they should be interpreted as a watercolor sketch of these relationships.

Multiple Representation Use and Algebra Classroom Variables

As the *Summary Report* noted, teachers who used multiple representations were more likely to allot more time for the topic and cover more material more deeply than other teachers. Tables 1-3 provide further evidence of this -- MRT's covered more topics in complex numbers. As might be expected, MRT's also had more reasons for covering the material and fewer reasons for NOT covering the material (Tables 4-17). Table 18 confirms results in the *Summary Report* by demonstrating that MRT cover a topic from complex numbers more deeply. This is consistent with the previously noted results that MRT's spend more time on the material. Tables 19-25 replicate the same results for teaching logarithms: MRT's allot more time, cover more material, cover the material more deeply, and have more reasons for covering the material.

Multiple Representation Use and Teacher Variables

There was a hint in the Summary Report that MRT's were better prepared professionally than non-MRT's. Tables 26 and 27 suggest that MRT's call on more students in a class period and spend more class time presenting new material than non-MRT's. The results in Table 28 appear to be random. Tables 29-41 illustrate statistically significantly relationships between MRT's and teachers' objectives and sources of information. Overall, the only pattern that begins to emerge is that MRT's are more likely to see a balanced variety of objectives and use a balanced variety of sources of information. This suggests that MRT's take a more reasoned, balanced approach to their teaching than non-MRT's. Interestingly, non-MRT's never use many sources of information.

MRT's also are more likely to divide the class into smaller groups (Table 42). Non-MRT's were more likely to assign the same homework to all students and they were more likely to blame the lack of student progress on the students themselves (Tables 43-46). The results given in Tables 47-49 do not fit any overall framework other than MRT's appear to be more reasonable and less dogmatic in their approaches than non-MRT's.

Multiple Representation Use and School Variables

As noted in the *Summary Report*, teachers' use of multiple representations are not significantly related to any school variables except the two listed in Tables 50 and 51 – school days per year and type of overall curriculum. Because of the lack of supporting variables, these results are attributed to chance.

Discussion

As noted above, these results fill in the picture of multiple representation use painted by the *Summary Report*, although the paints used are watercolors rather than oils. The results presented in the *Technical Appendix* suggest that teachers who use multiple representations cover more topics more deeply than teachers who do not use multiple representations. As one might expect, MRT's have more reasons for covering the material than non-MRT's and MRT's have fewer reasons for not covering topics. Also, these data hint that MRT's are more likely to avail themselves of different information sources than non-MRT's and they are more balanced in their views of mathematics, mathematics teaching, and mathematics learning. Of course, confirming these suggestions and hints would be an appropriate topic for further research.

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 1

TABLE OF COMPLEX USED BY ATTCXNN

COMPLEX USED ATTCXNN(83.TAUGHT NEW GRA[illegible] COMPLX NUM)

FREQUENCY PERCENT ROW PCT COL PCT	COVERED	NOT COVERED	TOTAL
USED <= 1	14 12.28 53.85 15.38	12 10.53 46.15 52.17	26 22.81
1 < USED <= 2	26 22.81 89.66 28.57	3 2.63 10.34 13.04	29 25.44
2 < USED	51 44.74 86.44 56.04	8 7.02 13.56 34.78	59 51.75
TOTAL	91 79.82	23 20.18	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY ATTCXNN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	14.239	0.001
LIKELIHOOD RATIO CHI-SQUARE	2	12.632	0.002
MANTEL-HAENSZEL CHI-SQUARE	1	9.266	0.002
PHI		0.353	
CONTINGENCY COEFFICIENT		0.333	
CRAMER'S V		0.353	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 2

TABLE OF COMPLEX USED BY ATTPOLN

COMPLEX USED ATTPOLN(88.TAUGHT NEW POLAR COORD COMP NUM)

FREQUENCY PERCENT ROW PCT COL PCT	COVERED	NOT COVERED	TOTAL
USED <= 1	11 9.91 42.31 14.47	15 13.51 57.69 42.86	26 23.42
1 < USED <= 2	22 19.82 75.86 28.95	7 6.31 24.14 20.00	29 26.13
2 < USED	43 38.74 76.79 56.58	13 11.71 23.21 37.14	56 50.45
TOTAL	76 68.47	35 31.53	111 100.00

FREQUENCY MISSING = 10

STATISTICS FOR TABLE OF COMPLEX USED BY ATTPOLN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	10.771	0.005
LIKELIHOOD RATIO CHI-SQUARE	2	10.202	0.006
MANTEL-HAENSZEL CHI-SQUARE	1	8.158	0.004
PHI		0.312	
CONTINGENCY COEFFICIENT		0.297	
CRAMER'S V		0.312	

EFFECTIVE SAMPLE SIZE = 111
FREQUENCY MISSING = 10

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 3

TABLE OF COMPLEX USED BY ATTDEMN

COMPLEX USED ATTDEMN(93.TAUGHT NEW DEM^ VRE'S THRM)

FREQUENCY / PERCENT / ROW PCT / COL PCT	COVERED	NOT COVERED	TOTAL
USED <= 1	12 10.53 46.15 15.00	14 12.28 53.85 41.18	26 22.81
1 < USED <= 2	24 21.05 82.76 30.00	5 4.39 17.24 14.71	29 25.44
2 < USED	44 38.60 74.58 55.00	15 13.16 25.42 44.12	59 51.75
TOTAL	80 70.18	34 29.82	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY ATTDEMN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	9.908	0.007
LIKELIHOOD RATIO CHI-SQUARE	2	9.485	0.009
MANTEL-HAENSZEL CHI-SQUARE	1	4.908	0.027
PHI		0.295	
CONTINGENCY COEFFICIENT		0.283	
CRAMER'S V		0.295	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 4

TABLE OF COMPLEX FREQ BY XPOSRTS

COMPLEX FREQ   XPOSRTS - NUMBER OF POSITIVE REASONS FOR TEACHING COMPLEX ROOTS

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 2 POSITIVE RE	3 <= REAS <= 4	4 <= REASONS	TOTAL
FREQ <= 1	30 24.79 38.96 78.95	28 23.14 36.36 62.22	19 15.70 24.68 50.00	77 63.64
FREQ >= 1	8 6.61 18.18 21.05	17 14.05 38.64 37.78	19 15.70 43.18 50.00	44 36.36
TOTAL	38 31.40	45 37.19	38 31.40	121 100.00

STATISTICS FOR TABLE OF COMPLEX FREQ BY XPOSRTS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.942	0.031
LIKELIHOOD RATIO CHI-SQUARE	2	7.167	0.028
MANTEL-HAENSZEL CHI-SQUARE	1	6.880	0.009
PHI		0.240	
CONTINGENCY COEFFICIENT		0.233	
CRAMER'S V		0.240	

SAMPLE SIZE = 121

CONTENT REP.ESENTATION: TECHNICAL APPENDIX
TABLE 5

TABLE OF COMPLEX FREQ BY XNOTRTS

COMPLEX FREQ XNOTRTS - NUMBER OF REASONS NOT MARKED FOR TEACHING COMPLEX ROOTS

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 4 REASONS NOT	5 <= REAS <= 6	7 <= REASONS	TOTAL
FREQ <= 1	19	29	29	77
	15.70	23.97	23.97	63.64
	24.68	37.66	37.66	
	50.00	63.04	78.38	
FREQ >= 1	19	17	8	44
	15.70	14.05	6.61	36.36
	43.18	38.64	18.18	
	50.00	36.96	21.62	
TOTAL	38	46	37	121
	31.40	38.02	30.58	100.00

STATISTICS FOR TABLE OF COMPLEX FREQ BY XNOTRTS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.535	0.038
LIKELIHOOD RATIO CHI-SQUARE	2	6.711	0.035
MANTEL-HAENSZEL CHI-SQUARE	1	5.887	0.015
PHI		0.232	
CONTINGENCY COEFFICIENT		0.226	
CRAMER'S V		0.232	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 6

TABLE OF COMPLEX FREQ BY XPOSCXN

COMPLEX FREQ   XPOSCXN - NUMBER OF POSITIVE REASONS MARKED FOR TEACHING GRAPHING COMPLEX NUMBERS

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 1 POSITIVE RE	2 <= REAS <= 3	4 <= REASONS	TOTAL
FREQ <= 1	28 23.14 36.36 80.00	26 21.49 33.77 65.00	23 19.01 29.87 50.00	77 63.64
FREQ >= 1	7 5.79 15.91 20.00	14 11.57 31.82 35.00	23 19.01 52.27 50.00	44 36.36
TOTAL	35 28.93	40 33.06	46 38.02	121 100.00

STATISTICS FOR TABLE OF COMPLEX FREQ BY XPOSCXN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	7.779	0.020
LIKELIHOOD RATIO CHI-SQUARE	2	8.033	0.018
MANTEL-HAENSZEL CHI-SQUARE	1	7.714	0.005
PHI		0.254	
CONTINGENCY COEFFICIENT		0.246	
CRAMER'S V		0.254	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 7
TABLE OF COMPLEX USED BY XPOSCXN

COMPLEX USED — XPOSCXN - NUMBER OF REASONS MARKED FOR TEACHING GRAPHING COMPLEX NUMBERS

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 1 POSITIVE RE	2 <= REAS <= 3	4 <= REASONS	TOTAL
USED <= 1	19 15.70 57.58 54.29	9 7.44 27.27 22.50	5 4.13 15.15 10.87	33 27.27
1 < USED <= 2	4 3.31 13.79 11.43	14 11.57 48.28 35.00	11 9.09 37.93 23.91	29 23.97
2 USED	12 9.92 20.34 34.29	17 14.05 28.81 42.50	30 24.79 50.85 65.22	59 48.76
TOTAL	35 28.93	40 33.06	46 38.02	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XPOSCXN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	22.945	0.000
LIKELIHOOD RATIO CHI-SQUARE	4	22.449	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	15.246	0.000
PHI		0.435	
CONTINGENCY COEFFICIENT		0.399	
CRAMER'S V		0.308	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 8

TABLE OF COMPLEX FREQ BY XNOTCXN

COMPLEX FREQ   XNOTCXN - NUMBER OF REASONS NOT MARKED FOR TEACHING GRAPHING COMPLEX NUMBERS

FREQUENCY PERCENT ROW PCT COL PCT	<= 5 REASONS NOT	6 <= REAS <= 7	8 <= REASONS	TOTAL
FREQ <= 1	23 19.01 29.87 50.00	35 28.93 45.45 71.43	19 15.70 24.68 73.08	77 63.64
FREQ >= 1	23 19.01 52.27 50.00	14 11.57 31.82 28.57	7 5.79 15.91 26.92	44 36.36
TOTAL	46 38.02	49 40.50	26 21.49	121 100.00

STATISTICS FOR TABLE OF COMPLEX FREQ BY XNOTCXN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	5.984	0.050
LIKELIHOOD RATIO CHI-SQUARE	2	5.937	0.051
MANTEL-HAENSZEL CHI-SQUARE	1	5.766	0.016
PHI		0.222	
CONTINGENCY COEFFICIENT		0.217	
CRAMER'S V		0.222	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 9

TABLE OF COMPLEX USED BY XNOTCXN

COMPLEX USED — XNOTCXN - NUMBER OF REASONS NOT MARKED FOR TEACHING GRAPHING COMPLEX NUMBERS

FREQUENCY PERCENT ROW PCT COL PCT	<= 5 REASONS NOT	6 <= REA <= 7	8 <= REASONS	TOTAL
USED <= 1	5 4.13 15.15 10.87	16 13.22 48.48 32.65	12 9.92 36.36 46.15	33 27.27
1 < USED <= 2	11 9.09 37.93 23.91	15 12.40 51.72 30.61	3 2.48 10.34 11.54	29 23.97
2 < USED	30 24.79 50.85 65.22	18 14.88 30.51 36.73	11 9.09 18.64 42.31	59 48.76
TOTAL	46 38.02	49 40.50	26 21.49	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XNOTCXN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	15.266	0.004
LIKELIHOOD RATIO CHI-SQUARE	4	16.226	0.003
MANTEL-HAENSZEL CHI-SQUARE	1	11.136	0.001
PHI		0.355	
CONTINGENCY COEFFICIENT		0.335	
CRAMER'S V		0.251	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 10

TABLE OF COMPLEX USED BY XPOSPOL

COMPLEX USED — XPOSPOL - NUMBER OF POSITIVE REASONS MARKED FOR TEACHING COMPLEX NUMBERS ON POLAR COORDS

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 1 POSITIVE REASONS	2 <= REASONS <= 3	4 <= REASONS	TOTAL
USED <= 1	22 18.18 66.67 44.90	6 4.96 18.18 17.65	5 4.13 15.15 13.16	33 27.27
1 < USED <= 2	8 6.61 27.59 16.33	11 9.09 37.93 32.35	10 8.26 34.48 26.32	29 23.97
2 < USED	19 15.70 32.20 38.78	17 14.05 28.81 50.00	23 19.01 38.98 60.53	59 48.76
TOTAL	49 40.50	34 28.10	38 31.40	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XPOSPOL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	13.882	0.008
LIKELIHOOD RATIO CHI-SQUARE	4	13.840	0.008
MANTEL-HAENSZEL CHI-SQUARE	1	8.814	0.003
PHI		0.339	
CONTINGENCY COEFFICIENT		0.321	
CRAMER'S V		0.240	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 11

TABLE OF COMPLEX USED BY XNEGPOL

COMPLEX USED — XNEGPOL - NUMBER OF NEGATIVE REASONS MARKED FOR NOT TEACHING COMPLEX NUMBERS ON POLAR COORDS

FREQUENCY / PERCENT / ROW PCT / COL PCT	0 NEGATIVE REASO	0 < NEGATIVE REA	TOTAL
USED <= 1	19 15.70 57.58 20.00	14 11.57 42.42 53.85	33 27.27
1 < USED <= 2	27 22.31 93.10 28.42	2 1.65 6.90 7.69	29 23.97
2 < USED	49 40.50 83.05 51.58	10 8.26 16.95 38.46	59 48.76
TOTAL	95 78.51	26 21.49	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XNEGPOL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	12.954	0.002
LIKELIHOOD RATIO CHI-SQUARE	2	12.682	0.002
MANTEL-HAENSZEL CHI-SQUARE	1	6.252	0.012
PHI		0.327	
CONTINGENCY COEFFICIENT		0.311	
CRAMER'S V		0.327	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 12

TABLE OF COMPLEX USED BY XNEGDEM

COMPLEX USED — XNEGDEM - NUMBER OF NEGATIVE REASONS MARKED FOR NOT TEACHING DEMOIVRE'S THEOREM

FREQUENCY / PERCENT / ROW PCT / COL PCT	0 NEGATIVE REASO	0 < NEGATIVE REA	TOTAL
USED <= 1	19 15.70 57.58 20.88	14 11.57 42.42 46.67	33 27.27
1 < USED <= 2	26 21.49 89.66 28.57	3 2.48 10.34 10.00	29 23.97
2 < USED	46 38.02 77.97 50.55	13 10.74 22.03 43.33	59 48.76
TOTAL	91 75.21	30 24.79	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XNEGDEM

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	8.989	0.011
LIKELIHOOD RATIO CHI-SQUARE	2	9.030	0.011
MANTEL-HAENSZEL CHI-SQUARE	1	3.413	0.065
PHI		0.273	
CONTINGENCY COEFFICIENT		0.263	
CRAMER'S V		0.273	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 13

TABLE OF COMPLEX USED BY XNOTDEM

COMPLEX USED — XNOTDEM - NUMBER OF REASONS NOT MARKED FOR TEACHING DEMOIVRE'S THEOREM

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 5 REASONS NOT	6 <= REAS <= 7	8 <= REASONS	TOTAL
USED <= 1	4 3.31 12.12 10.00	15 12.40 45.45 31.25	14 11.57 42.42 42.42	33 27.27
1 < USED <= 2	11 9.09 37.93 27.50	14 11.57 48.28 29.17	4 3.31 13.79 12.12	29 23.97
2 < USED	25 20.66 42.37 62.50	19 15.70 32.20 39.58	15 12.40 25.42 45.45	59 48.76
TOTAL	40 33.06	48 39.67	33 27.27	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XNOTDEM

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	12.565	0.014
LIKELIHOOD RATIO CHI-SQUARE	4	13.881	0.008
MANTEL-HAENSZEL CHI-SQUARE	1	7.881	0.005
PHI		0.322	
CONTINGENCY COEFFICIENT		0.307	
CRAMER'S V		0.228	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 14

TABLE OF COMPLEX FREQ BY XTOTPOS

COMPLEX FREQ XTOTPOS - TOTAL NUMBER OF POSITIVE REASONS MARKED

REQUENCY / PERCENT / ROW PCT / COL PCT	<= 7 POSITIVE RE	8 <= REAS <= 14	14 <= REASONS	TOTAL
FREQ <= 1	33 27.27 42.86 80.49	22 18.18 28.57 56.41	22 18.18 28.57 53.66	77 63.64
FREQ >= 1	8 6.61 18.18 19.51	17 14.05 38.64 43.59	19 15.70 43.18 46.34	44 36.36
TOTAL	41 33.88	39 32.23	41 33.88	121 100.00

STATISTICS FOR TABLE OF COMPLEX FREQ BY XTOTPOS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	7.675	0.022
LIKELIHOOD RATIO CHI-SQUARE	2	8.113	0.017
MANTEL-HAENSZEL CHI-SQUARE	1	6.533	0.011
PHI		0.252	
CONTINGENCY COEFFICIENT		0.244	
CRAMER'S V		0.252	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL PPENDIX

TABLE 15

TABLE OF COMPLEX USED BY XTOTPOS

COMPLEX USED XTOTPOS - TOTAL NUMBER OF POSITIVE REASONS MARKED

FREQUENCY PERCENT ROW PCT COL PCT	<= 7 POS ITIVE RE	8 <= REA S <= 14	14 <= RE ASONS	TOTAL
USED <= 1	22 18.18 66.67 53.66	6 4.96 18.18 15.38	5 4.13 15.15 12.20	33 27.27
1 < USED <= 2	5 4.13 17.24 12.20	13 10.74 44.83 33.33	11 9.09 37.93 26.83	29 23.97
2 < USED	14 11.57 23.73 34.15	20 16.53 33.90 51.28	25 20.66 42.37 60.98	59 48.76
TOTAL	41 33.88	39 32.23	41 33.88	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XTOTPOS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	22.945	0.000
LIKELIHOOD RATIO CHI-SQUARE	4	22.380	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	13.615	0.000
PHI		0.435	
CONTINGENCY COEFFICIENT		0.399	
CRAMER'S V		0.308	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 16

TABLE OF COMPLEX USED BY XTOTNEG

COMPLEX USED — XTOTNEG - TOTAL NUMBER OF NEGATIVE REASONS MARKED

FREQUENCY PERCENT ROW PCT COL PCT	0 NEGATIVE REASO	0 < NEGATIVE REA	TOTAL
USED <= 1	14 11.57 42.42 16.87	19 15.70 57.58 50.00	33 27.27
1 < USED <= 2	26 21.49 89.66 31.33	3 2.48 10.34 7.89	29 23.97
2 < USED	43 35.54 72.88 51.81	16 13.22 27.12 42.11	59 48.76
TOTAL	83 68.60	38 31.40	121 100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XTOTNEG

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	16.966	0.000
LIKELIHOOD RATIO CHI-SQUARE	2	17.356	0.000
MANTEL-HAENSZEL CHI-SQUARE	1	6.641	0.010
PHI		0.374	
CONTINGENCY COEFFICIENT		0.351	
CRAMER'S V		0.374	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 17
TABLE OF COMPLEX USED BY XTOTNOT

COMPLEX USED XTOTNOT - TOTAL NUMBER OF REASONS NOT MARKED

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 20 REASONS NO	21= < REAS <= 28	29 <= REASONS	TOTAL
USED <= 1	5	15	13	33
	4.13	12.40	10.74	27.27
	15.15	45.45	39.39	
	12.82	26.32	52.00	
1 < USED <= 2	10	16	3	29
	8.26	13.22	2.48	23.97
	34.48	55.17	10.34	
	25.64	28.07	12.00	
2 < USED	24	26	9	59
	19.83	21.49	7.44	48.76
	40.68	44.07	15.25	
	61.54	45.61	36.00	
TOTAL	39	57	25	121
	32.23	47.11	20.66	100.00

STATISTICS FOR TABLE OF COMPLEX USED BY XTOTNOT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	12.807	0.012
LIKELIHOOD RATIO CHI-SQUARE	4	12.658	0.013
MANTEL-HAENSZEL CHI-SQUARE	1	7.888	0.005
PHI		0.325	
CONTINGENCY COEFFICIENT		0.309	
CRAMER'S V		0.230	

SAMPLE SIZE = 121

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 18

TABLE OF COMPLEX USED BY APRCOS

COMPLEX USED APRCOS(217.PRESENTED R COSINE THETA)

FREQUENCY PERCENT ROW PCT COL PCT	GAVE FOR MAL PROO	STATED W DERIV	STATED-N O DERIV	DID NOT COVER	TOTAL
USED <= 1	2 1.71 6.90 4.35	6 5.13 20.69 27.27	5 4.27 17.24 31.25	16 13.68 55.17 48.48	29 24.79
1 < USED <= 2	15 12.82 51.72 32.61	5 4.27 17.24 22.73	4 3.42 13.79 25.00	5 4.27 17.24 15.15	29 24.79
2 < USED	29 24.79 49.15 63.04	11 9.40 18.64 50.00	7 5.98 11.86 43.75	12 10.26 20.34 36.36	59 50.43
TOTAL	46 39.32	22 18.80	16 13.68	33 28.21	117 100.00

FREQUENCY MISSING = 4

STATISTICS FOR TABLE OF COMPLEX USED BY APRCOS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	20.852	0.002
LIKELIHOOD RATIO CHI-SQUARE	6	23.409	0.001
MANTEL-HAENSZEL CHI-SQUARE	1	14.083	0.000
PHI		0.422	
CONTINGENCY COEFFICIENT		0.389	
CRAMER'S V		0.299	

EFFECTIVE SAMPLE SIZE = 117
FREQUENCY MISSING = 4

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 19

TABLE OF LOG FREQ BY ATTGRLN

LOG FREQ   ATTGRLN(43.TAUGHT NEW GRAPHING LOG FUNCT)

FREQUENCY PERCENT ROW PCT COL PCT	COVERED	NOT COVERED	TOTAL
FREQ <= 1	59 51.30 71.08 67.05	24 20.87 28.92 88.89	83 72.[illegible]
FREQ >= 1	29 25.22 90.63 32.95	3 2.61 9.38 11.11	32 27.83
TOTAL	88 76.52	27 23.48	115 100.00

FREQUENCY MISSING = 1

STATISTICS FOR TABLE OF LOG FREQ BY ATTGRLN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	1	4.909	0.027
LIKELIHOOD RATIO CHI-SQUARE	1	5.604	0.018
CONTINUITY ADJ. CHI-SQUARE	1	3.881	0.049
MANTEL-HAENSZEL CHI-SQUARE	1	4.866	0.027
FISHER'S EXACT TEST (1-TAIL)			0.020
(2-TAIL)			0.028
PHI		-0.207	
CONTINGENCY COEFFICIENT		0.202	
CRAMER'S V		-0.207	

EFFECTIVE SAMPLE SIZE = 115
FREQUENCY MISSING = 1

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 20

TABLE OF LOG USED BY ATTGRLN

LOG USED ATTGRLN(43.TAUGHT NEW GRAPHING LOG FJNCT)

FREQUENCY PERCENT ROW PCT COL PCT	COVERED	NOT COVERED	TOTAL
USED <= 1	20 17.39 58.82 22.73	14 12.17 41.18 51.85	34 29.57
1 < USED <= 2	44 38.26 80.00 50.00	11 9.57 20 00 40.74	55 47.83
2 < USED	24 20.87 92.31 27.27	2 1.74 7.69 7.41	26 22.61
TOTAL	88 76.52	27 23.48	115 100.00

FREQUENCY MISSING = 1

STATISTICS FOR TABLE OF LOG USED BY ATTGRLN

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	9.904	0.007
LIKELIHOOD RATIO CHI-SQUARE	2	10.132	0.006
MANTEL-HAENSZEL CHI-SQUARE	1	9.510	0.002
PHI		0.293	
CONTINGENCY COEFFICIENT		0.282	
CRAMER'S V		0.293	

EFFECTIVE SAMPLE SIZE = 115
FREQUENCY MISSING = 1

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 21

TABLE OF LOG FREQ BY XPOSGRL

LOG FREQ   XPOSGRL - NUMBER OF POSITIVE REASONS MARKED FOR TEACHING GRAPHING LOG FUNCTIONS

FREQUENCY PERCENT ROW PCT COL PCT	<= 1 POSITIVE RE	2 <= REAS <= 3	4 <= REASONS	TOTAL
FREQ <= 1	28	24	32	84
	24.14	20.69	27.59	72.41
	33.33	28.57	38.10	
	87.50	60.00	72.73	
FREQ >= 1	4	16	12	32
	3.45	13.79	10.34	27.59
	12.50	50.00	37.50	
	12.50	40.00	27.27	
TOTAL	32	40	44	116
	27.59	34.48	37.93	100.00

STATISTIC FOR TABLE OF LOG FREQ BY XPOSGRL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.734	0.034
LIKELIHOOD RATIO CHI-SQUARE	2	7.131	0.028
MANTEL-HAENSZEL CHI-SQUARE	1	1.460	0.227
PHI		0.241	
CONTINGENCY COEFFICIENT		0.234	
CRAMER'S V		0.241	

SAMPLE SIZE = 116

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 22

TABLE OF LOG FREQ BY XNEGGRL

LOG FREQ   XNEGGRL - NUMBER OF NEGATIVE REASONS MARKED FOR NOT TEACHING GRAPHING LOG FUNCTIONS

FREQUENCY PERCENT ROW PCT COL PCT	0 NEGATI VE REASO	0 < NEGA TIVE REA	TOTAL
FREQ <= 1	61 52.59 72.62 66.30	23 19.83 27.38 95.83	84 72.41
FREQ >= 1	31 26.72 96.88 33.70	1 0.86 3.13 4.17	32 27.59
TOTAL	92 79.31	24 20.69	116 100.00

STATISTICS FOR TABLE OF LOG FREQ BY XNEGGRL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	1	8.309	0.004
LIKELIHOOD RATIO CHI-SQUARE	1	10.759	0.001
CONTINUITY ADJ. CHI-SQUARE	1	6.896	0.009
MANTEL-HAENSZEL CHI-SQUARE	1	8.237	0.004
FISHER'S EXACT TEST (1-TAIL)			0.002
(2-TAIL)			0.004
PHI		-0.268	
CONTINGENCY COEFFICIENT		0.259	
CRAMER'S V		-0.268	

SAMPLE SIZE = 116

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 23

TABLE OF LOG USED BY XNEGGRL

LOG USED — XNEGGRL - TOTAL NUMBER OF REASONS MARKED FOR NOT TEACHING GRAPHING LOG FUNCTIONS

FREQUENCY PERCENT ROW PCT COL PCT	0 NEGATIVE REASO	0 < NEGATIVE REA	TOTAL
USED <= 1	22 18.97 64.71 23.91	12 10.34 35.29 50.00	34 29.31
1 < USED <= 2	45 38.79 80.36 48.91	11 9.48 19.64 45.83	56 48.28
2 < USED	25 21.55 96.15 27.17	1 0.86 3.85 4.17	26 22.41
TOTAL	92 79.31	24 20.69	116 100.00

STATISTICS FOR TABLE OF LOG USED BY XNEGGRL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	8.952	0.011
LIKELIHOOD RATIO CHI-SQUARE	2	10.165	0.006
MANTEL-HAENSZEL CHI-SQUARE	1	8.875	0.003
PHI		0.278	
CONTINGENCY COEFFICIENT		0.268	
CRAMER'S V		0.278	

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 24

TABLE OF LOG USED BY XTOTNEG

LOG USED — XTOTNEG - TOTAL NUMBER OF NEGATIVE REASONS MARKED

FREQUENCY / PERCENT / ROW PCT / COL PCT	0 NEGATIVE REASO	0 < NEGATIVE REA	TOTAL
USED <= 1	18 15.52 52.94 25.00	16 13.79 47.06 36.36	34 29.31
1 < USED <= 2	31 26.72 55.36 43.06	25 21.55 44.64 56.82	56 48.28
2 < USED	23 19.83 88.46 31.94	3 2.59 11.54 6.82	26 22.41
TOTAL	72 62.07	44 37.93	116 100.00

STATISTICS FOR TABLE OF LOG USED BY XTOTNEG

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	9.967	0.007
LIKELIHOOD RATIO CHI-SQUARE	2	11.383	0.003
MANTEL-HAENSZEL CHI-SQUARE	1	7.034	0.008
PHI		0.293	
CONTINGENCY COEFFICIENT		0.281	
CRAMER'S V		0.293	

SAMPLE SIZE = 116

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 25

TABLE OF LOG FREQ BY AEXLOGB

LOG FREQ AEXLOGB(203.EXPECT LOG BASE B OF X)

FREQUENCY / PERCENT / ROW PCT / COL PCT	PROVE AND APPLY	DERIVE AND APPLY	RECALL AND APPLY	WHEN GIVEN,APPLY	NOT DISCUSSED	TOTAL
FREQ <= 1	2	15	34	8	16	75
	1.89	14.15	32.08	7.55	15.09	70.75
	2.67	20.00	45.33	10.67	21.33	
	25.00	78.95	69.39	66.67	88.89	
FREQ >= 1	6	4	15	4	2	31
	5.66	3.77	14.15	3.77	1.89	29.25
	19.35	12.90	48.39	12.90	6.45	
	75.00	21.05	30.61	33.33	11.11	
TOTAL	8	19	49	12	18	106
	7.55	17.92	46.23	11.32	16.98	100.00

FREQUENCY MISSING = 10

STATISTICS FOR TABLE OF LOG FREQ BY AEXLOGB

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	11.712	0.020
LIKELIHOOD RATIO CHI-SQUARE	4	11.366	0.023
MANTEL-HAENSZEL CHI-SQUARE	1	4.998	0.025
PHI		0.332	
CONTINGENCY COEFFICIENT		0.315	
CRAMER'S V		0.332	

EFFECTIVE SAMPLE SIZE = 106
FREQUENCY MISSING = 10

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 26

TABLE OF COMPLEX USED BY TQUEST

COMPLEX USED   TQUEST - DIFFERENT STUDENTS QUESTIONED

FREQUENCY / PERCENT / ROW PCT / COL PCT	<= 25% 1	25% < N <= 50% 2	50% < N <= 75% 3	N > 75% 4	TOTAL
USED <= 1	14 12.28 43.75 45.16	5 4.39 15.63 16.67	10 8.77 31.25 40.00	3 2.63 9.38 10.71	32 28.07
1 < USED <= 2	6 5.26 25.00 19.35	4 3.51 16.67 13.33	8 7.02 33.33 32.00	6 5.26 25.00 21.43	24 21.05
2 < USED	11 9.65 18.97 35.48	21 18.42 36.21 70.00	7 6.14 12.07 28.00	19 16.67 32.76 67.86	58 50.88
TOTAL	31 27.19	30 26.32	25 21.93	28 24.56	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY TQUEST

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	6	18.963	0.004
LIKELIHOOD RATIO CHI-SQUARE	6	19.712	0.003
MANTEL-HAENSZEL CHI-SQUARE	1	3.904	0.048
PHI		0.408	
CONTINGENCY COEFFICIENT		0.378	
CRAMER'S V		0.288	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 27

TABLE OF COMPLEX USED BY TEXPLNT

COMPLEX USED FREQUENCY PERCENT ROW PCT COL PCT	TEXPLNT - MINUTES EXPLAINING NEW MATERIAL TYPICAL WEEK MINS < 100	100 <= MINS < 15	150 <= MINS	TOTAL
USED <= 1	11 9.65 34.38 33.33	16 14.04 50.00 39.02	5 4.39 15.63 12.50	32 28.07
1 < USED <= 2	9 7.89 37.50 27.27	6 5.26 25.00 14.63	9 7.89 37.50 22.50	24 21.05
2 < USED	13 11.40 22.41 39.39	19 16.67 32.76 46.34	26 22.81 44.83 65.00	58 50.88
TOTAL	33 28.95	41 35.96	40 35.09	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY TEXPLNT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	9.571	0.048
LIKELIHOOD RATIO CHI-SQUARE	4	10.266	0.036
MANTEL-HAENSZEL CHI-SQUARE	1	4.690	0.030
PHI		0.290	
CONTINGENCY COEFFICIENT		0.278	
CRAMER'S V		0.205	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7
SAMPLE SIZE = 116

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 28

TABLE OF LOG FREQ BY TNONEW

LOG FREQ TNONEW - NOT ANY DISCOVERIES FOR A LONG TIME

FREQUENCY / PERCENT / ROW PCT / COL PCT	STRONG DISAGREE 1	DISAGREE 2	UN-DECIDED 3	TOTAL
FREQ <= 1	25 23.15 32.47 89.29	45 41.67 58.44 63.38	7 6.48 9.09 77.78	77 71.30
FREQ >= 1	3 2.78 9.68 10.71	26 24.07 83.87 36.62	2 1.85 6.45 22.22	31 28.70
TOTAL	28 25.93	71 65.74	9 8.33	108 100.00

FREQUENCY MISSING = 8

STATISTICS FOR TABLE OF LOG FREQ BY TNONEW

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.787	0.034
LIKELIHOOD RATIO CHI-SQUARE	2	7.605	0.022
MANTEL-HAENSZEL CHI-SQUARE	1	2.853	0.091
PHI		0.251	
CONTINGENCY COEFFICIENT		0.243	
CRAMER'S V		0.251	

EFFECTIVE SAMPLE SIZE = 108
FREQUENCY MISSING = 8

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 29

TABLE OF LOG USED BY ROBJINT

LOG USED ROBJINT(3.OBJECTIVE..INTEREST IN MATHEMATICS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	RELATIVELY MORE	EQUAL EMPHASIS	RELATIVELY LESS	TOTAL
USED <= 1	12	16	4	32
	10.53	14.04	3.51	28.07
	37.50	50.00	12.50	
	31.58	27.59	22.22	
1 < USED <= 2	15	35	6	56
	13.16	30.70	5.26	49.12
	26.79	62.50	10.71	
	39.47	60.34	33.33	
2 < USED	11	7	8	26
	9.65	6.14	7.02	22.81
	42.31	26.92	30.77	
	28.95	12.07	44.44	
TOTAL	38	58	18	114
	33.33	50.88	15.79	100.00

FREQUENCY MISSING 2

STATISTICS FOR TABLE OF LOG USED BY ROBJINT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	10.767	0.029
LIKELIHOOD RATIO CHI-SQUARE	4	10.60	0.031
MANTEL-HAENSZEL CHI-SQUARE	1	0.582	0.446
PHI		0.307	
CONTINGENCY COEFFICIENT		0.294	
CRAMER'S V		0.217	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 30

TABLE OF LOG USED BY ROBJLIF

LOG USED FREQUENCY PERCENT ROW PCT COL PCT	ROBJLIF(6.OBJECTIVE..AWARENESS OF MATH IN LIFE) RELATIVELY MORE	EQUAL EMPHASIS	RELATIVELY LESS	TOTAL
USED <= 1	16 14.04 50.00 48.48	10 8.77 31.25 16.67	6 5.26 18.75 28.57	32 28.07
1 < USED <= 2	12 10.53 21.43 36.36	33 28.95 58.93 55 00	11 9.65 19.64 52.38	56 49.12
2 < USED	5 4.39 19.23 15.15	17 14.91 65.38 28.33	4 3.51 15.38 19.05	26 22.81
TOTAL	33 28.95	60 52.63	21 18.42	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG USED BY ROBJLIF

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	11.023	0.026
LIKELIHOOD RATIO CHI-SQUARE	4	10.776	0.029
MANTEL-HAENSZEL CHI-SQUARE	1	2.601	0.107
PHI		0.311	
CONTINGENCY COEFFICIENT		0.297	
CRAMER'S V		0.220	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 31

TABLE OF LOG USEO BY ROBJCOM

LOG USEO / FREQUENCY / PERCENT / ROW PCT / COL PCT	ROBJCOM(7.OBJECTIVE..COMPUTATION SPEEO ACCURACY) RELATIVELY MORE	EQUAL EMPHASIS	RELATIVELY LESS	TOTAL
USEO <= 1	13 11.40 40.63 44.83	12 10.53 37.50 20.34	7 6.14 21.88 26.92	32 28.07
1 < USEO <= 2	14 12.28 25.00 48.28	28 24.56 50.00 47.46	14 12.28 25.00 53.85	56 49.12
2 < USEO	2 1.75 7.69 6.90	19 16.67 73.08 32.20	5 4.39 19.23 19.23	26 22.81
TOTAL	29 25.44	59 51.75	26 22.81	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG USEO BY ROBJCOM

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	9.974	0.041
LIKELIHOOD RATIO CHI-SQUARE	4	10.628	0.031
MANTEL-HAENSZEL CHI-SQUARE	1	2.789	0.095
PHI		0.296	
CONTINGENCY COEFFICIENT		0.284	
CRAMER'S V		0.209	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 32

TABLE OF COMPLEX USED BY ROBJSCI

COMPLEX USED ROBJSCI(8.OBJECTIVE..AWARE OF MATH IN SCIENCE)

FREQUENCY / PERCENT / ROW PCT / COL PCT	RELATIVELY MORE	EQUAL EMPHASIS	RELATIVELY LESS	TOTAL
USED <= 1	11	19	2	32
	9.17	15.83	1.67	26.67
	34.38	59.38	6.25	
	28.95	30.65	10.00	
1 < USED <= 2	8	11	10	29
	6.67	9.17	8.33	24.17
	27.59	37.93	34.48	
	21.05	17.74	50.00	
2 < USED	19	32	8	59
	15.83	26.67	6.67	49.17
	32.20	54.24	13.56	
	50.00	51.61	40.00	
TOTAL	38	62	20	120
	31.67	51.67	16.67	100.00

FREQUENCY MISSING = 1

STATISTICS FOR TABLE OF COMPLEX USED BY ROBJSCI

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	9.683	0.046
LIKELIHOOD RATIO CHI-SQUARE	4	9.146	0.058
MANTEL-HAENSZEL CHI-SQUARE	1	0.106	0.744
PHI		0.284	
CONTINGENCY COEFFICIENT		0.273	
CRAMER'S V		0.201	

EFFECTIVE SAMPLE SIZE = 120
FREQUENCY MISSING = 1

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 33

TABLE OF LOG FREQ BY ROBJSCI

LOG FREQ FREQUENCY PERCENT ROW PCT COL PCT	ROBJSCI(8.OBJECTIVE..AWARE OF MATH IN SCIENCE) RELATIVELY MORE	EQUAL EMPHASIS	RELATIVELY LESS	TOTAL
FREQ <= 1	17 14.91 20.73 48.57	52 45.61 63.41 82.54	13 11.40 15.85 81.25	82 71.93
FREQ >= 1	18 15.79 56.25 51.43	11 9.65 34.38 17.46	3 2.63 9.38 18.75	32 28.07
TOTAL	35 30.70	63 55.26	16 14.04	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG FREQ BY ROBJSCI

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	13.659	0.001
LIKELIHOOD RATIO CHI-SQUARE	2	13.058	0.001
MANTEL-HAENSZEL CHI-SQUARE	1	9.591	0.002
PHI		0.346	
CONTINGENCY COEFFICIENT		0.327	
CRAMER'S V		0.346	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 34

TABLE OF LOG FREQ BY RSISYLG

LOG FREQ   RSISYLG(10B.GOALS SOURCE..SYLLABUS)

FREQUENCY PERCENT ROW PCT COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
FREQ <= 1	27 25.23 36.00 90.00	37 34.58 49.33 61.67	11 10.28 14.67 64.71	75 70.09
FREQ >= 1	3 2.80 9.38 10.00	23 21.50 71.88 38.33	6 5.61 18.75 35.29	32 29.91
TOTAL	30 28.04	60 56.07	17 15.89	107 100.00

FREQUENCY MISSING = 9

STATISTICS FOR TABLE OF LOG FREQ BY RSISYLG

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	7.939	0.019
LIKELIHOOD RATIO CHI-SQUARE	2	9.095	0.011
MANTEL-HAENSZEL CHI-SQUARE	1	4.936	0.026
PHI		0.272	
CONTINGENCY COEFFICIENT		0.263	
CRAMER'S V		0.272	

EFFECTIVE SAMPLE SIZE = 107
FREQUENCY MISSING = 9

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 35
TABLE OF LOG USED BY RSIPROG

LOG USED RSIPROG(10F GOALS SOURCE..PROF MEETINGS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
USED <= 1	4 3.77 13.33 19.05	12 11.32 40.00 20.69	14 13.21 46.67 51.85	30 28.30
1 < USED <= 2	11 10.38 20.75 52.38	35 33.02 66.04 60.34	7 6.60 13.21 25.93	53 50.00
2 < USED	6 5.66 26.09 28.57	11 10.38 47.83 18.97	6 5.66 26.09 22.22	23 21.70
TOTAL	21 19.81	58 54.72	27 25.47	106 100.00

FREQUENCY MISSING = 10

STATISTICS FOR TABLE OF LOG USED BY RSIPROG

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	12.169	0.016
LIKELIHOOD RATIO CHI-SQUARE	4	11.885	0.018
MANTEL-HAENSZEL CHI-SQUARE	1	3.868	0.049
PHI		0.339	
CONTINGENCY COEFFICIENT		0.321	
CRAMER'S V		0.240	

EFFECTIVE SAMPLE SIZE = 106
FREQUENCY MISSING = 10

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 36

TABLE OF LOG FREQ BY RSIJRNP

LOG FREQ   RSIJRNP(11E.PRESENTATION SOURCE..JOURNALS, BOOKS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
FREQ <= 1	50 46.73 65.79 83.33	17 15.89 22.37 58.62	9 8.41 11.84 50.00	76 71.03
FREQ >= 1	10 9.35 32.26 16.67	12 11.21 38.71 41.38	9 8.41 29.03 50.00	31 28.97
TOTAL	60 56.07	29 27.10	18 16.82	107 100.00

FREQUENCY MISSING = 9

STATISTICS FOR TABLE OF LOG FREQ BY RSIJRNP

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	10.452	0.005
LIKELIHOOD RATIO CHI-SQUARE	2	10.450	0.005
MANTEL-HAENSZEL CHI-SQUARE	1	9.761	0.002
PHI		0.313	
CONTINGENCY COEFFICIENT		0.298	
CRAMER'S V		0.313	

EFFECTIVE SAMPLE SIZE = 107
FREQUENCY MISSING = 9

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 37

TABLE OF LOG FREQ BY RSIOTHP

LOG FREQ RSIOTHP(11G.PRESENTATION SOURCE..OTHER TEACHERS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
FREQ <= 1	49 46.23 65.33 79.03	22 20.75 29.33 66.67	4 3.77 5.33 36.36	75 70.75
FREQ >= 1	13 12.26 41.94 20.97	11 10.38 35.48 33.33	7 6.60 22.58 63.64	31 29.25
TOTAL	62 58.49	33 31.13	11 10.38	106 100.00

FREQUENCY MISSING = 10

STATISTICS FOR TABLE OF LOG FREQ BY RSIOTHP

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	8.607	0.014
LIKELIHOOD RATIO CHI-SQUARE	2	8.011	0.018
MANTEL-HAENSZEL CHI-SQUARE	1	7.851	0.005
PHI		0.285	
CONTINGENCY COEFFICIENT		0.274	
CRAMER'S V		0.285	

EFFECTIVE SAMPLE SIZE = 106
FREQUENCY MISSING = 10

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 38

TABLE OF COMPLEX USED BY RSITXTD

COMPLEX USED — RSITXTD(12A.DRILL SOURCE..TEXTBOOK)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
USED <= 1	23 20.18 76.67 38.98	3 2.63 10.00 8.82	4 3.51 13.33 19.05	30 26.32
1 < USED <= 2	10 8.77 37.04 16.95	7 6.14 25.93 20.59	10 8.77 37.04 47.62	27 23.68
2 < USED	26 22.81 45.61 44.07	24 21.05 42.11 70.59	7 6.14 12.28 33.33	57 50 00
TOTAL	59 51.75	34 29.82	21 18.42	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY RSITXTD

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	18.784	0.001
LIKELIHOOD RATIO CHI-SQUARE	4	18.558	0.001
MANTEL-HAENSZEL CHI-SQUARE	1	1.686	0.194
PHI		0.406	
CONTINGENCY COEFFICIENT		0.376	
CRAMER'S V		0.287	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 39

TABLE OF COMPLEX FREQ BY RSISYLA

COMPLEX FREQ   RSISYLA(13B.APPLICATIONS SOURCE..SYLLABUS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
FREQ <= 1	29 25.44 40.28 80.56	34 29.82 47.22 59.65	9 7.89 12.50 42.86	72 63.16
FREQ >= 1	7 6.14 16.67 19.44	23 20.18 54.76 40.35	12 10.53 28.57 57.14	42 36.84
TOTAL	36 31.58	57 50.00	21 18.42	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX FREQ BY RSISYLA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	8.704	0.013
LIKELIHOOD RATIO CHI-SQUARE	2	9.017	0.011
MANTEL-HAENSZEL CHI-SQUARE	1	8.578	0.003
PHI		0.276	
CONTINGENCY COEFFICIENT		0.266	
CRAMER'S V		0.276	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 40

TABLE OF COMPLEX USED BY RSISYLA

COMPLEX USED RSISYLA(13B.APPLICATIONS SOURCE..SYLLABUS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
USED <= 1	15 13.16 50.00 41.67	11 9.65 36.67 19.30	4 3.51 13.33 19.05	30 26.32
1 < USED <= 2	10 8.77 37.04 27.78	14 12.28 51.85 24.56	3 2.63 11.11 14.29	27 23.68
2 < USED	11 9.65 19.30 30.56	32 28.07 56.14 56.14	14 12.28 24.56 66.67	57 50.00
TOTAL	36 31.58	57 50.00	21 18.42	114 100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX USED BY RSISYLA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	10.087	0.039
LIKELIHOOD RATIO CHI-SQUARE	4	10.184	0.037
MANTEL-HAENSZEL CHI-SQUARE	1	7.849	0.005
PHI		0.297	
CONTINGENCY COEFFICIENT		0.285	
CRAMER'S V		0.210	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 41

TABLE OF LOG FREQ BY RSITXTA

LOG FREQ RSITXTA(13A.APPLICATIONS SOURCE..TEXTBOOK)

FREQUENCY / PERCENT / ROW PCT / COL PCT	NEVER USED	OCCASIONALLY	FREQUENTLY USED	TOTAL
FREQ <= 1	36	29	10	75
	33.64	27.10	9.35	70.09
	48.00	38.67	13.33	
	81.82	56.86	83.33	
FREQ >= 1	8	22	2	32
	7.48	20.56	1.87	29.91
	25.00	68.75	6.25	
	18.18	43.14	16.67	
TOTAL	44	51	12	107
	41.12	47.66	11.21	100.00

FREQUENCY MISSING = 9

STATISTICS FOR TABLE OF LOG FREQ BY RSITXTA

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	8.143	0.017
LIKELIHOOD RATIO CHI-SQUARE	2	8.280	0.016
MANTEL-HAENSZEL CHI-SQUARE	1	1.297	0.255
PHI		0.276	
CONTINGENCY COEFFICIENT		0.266	
CRAMER'S V		0.276	

EFFECTIVE SAMPLE SIZE = 107
FREQUENCY MISSING = 9

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 42

TABLE OF LOG USED BY RGRPWHL

LOG USED FREQUENCY PERCENT ROW PCT COL PCT	RGRPWHL(28.WHOLE CLASS WORKING AS A SINGLE GROUP) % < 60	60 <= % < 75	75 <= %	TOTAL
USED <= 1	6 5.26 18.75 20.00	4 3.51 12.50 12.50	22 19.30 68.75 42.31	32 28.07
1 < USED <= 2	17 14.91 30.36 56.67	16 14.04 28.57 50.00	23 20.18 41.07 44.23	56 49.12
2 < USED	7 6.14 26.92 23.33	12 10.53 46.15 37.50	7 6.14 26.92 13.46	26 22.81
TOTAL	30 26.32	32 28.07	52 45.61	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG USED BY RGRPWHL

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	12.845	0.012
LIKELIHOOD RATIO CHI-SQUARE	4	12.983	0.011
MANTEL-HAENSZEL CHI-SQUARE	1	2.032	0.154
PHI		0.336	
CONTINGENCY COEFFICIENT		0.318	
CRAMER'S V		0.237	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 43

TABLE OF LOG FREQ BY RHDNAPP

LOG FREQ RHDNAPP(42.SOME STUDENTS..NOT APPLICABLE)

FREQUENCY / PERCENT / ROW PCT / COL PCT	YES	NO	TOTAL
FREQ <= 1	57 50.00 69.51 66.28	25 21.93 30.49 89.29	82 71.93
FREQ >= 1	29 25.44 90.63 33.72	3 2.63 9.38 10.71	32 28.07
TOTAL	86 75.44	28 24.56	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG FREQ BY RHDNAPP

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	1	5.537	0.019
LIKELIHOOD RATIO CHI-SQUARE	1	6.339	0.012
CONTINUITY ADJ. CHI-SQUARE	1	4.457	0.035
MANTEL-HAENSZEL CHI-SQUARE	1	5.489	0.019
FISHER'S EXACT TEST (1-TAIL)			0.014
(2-TAIL)			0.028
PHI		-0.220	
CONTINGENCY COEFFICIENT		0.215	
CRAMER'S V		-0.220	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 44

TABLE OF LOG FREQ BY RPGINDF

LOG FREQ RPGINDF(45.PROGRESS..STUDENT INDIFFERENCE)

FREQUENCY / PERCENT / ROW PCT / COL PCT	IMPORTANT REASON	SOMEWHAT IMPORTANT	NOT IMPORTANT	TOTAL
FREQ <= 1	44	35	3	82
	38.60	30.70	2.63	71.93
	53.66	42.68	3.66	
	77.19	72.92	33.33	
FREQ >= 1	13	13	6	32
	11.40	11.40	5.26	28.07
	40.63	40.63	18.75	
	22.81	27.08	66.67	
TOTAL	57	48	9	114
	50.00	42.11	7.89	100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG FREQ BY RPGINDF

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	7.445	0.024
LIKELIHOOD RATIO CHI-SQUARE	2	6.604	0.037
MANTEL-HAENSZEL CHI-SQUARE	1	4.493	0.034
PHI		0.256	
CONTINGENCY COEFFICIENT		0.248	
CRAMER'S V		0.256	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 45

TABLE OF COMPLEX FREQ BY RPGABS

COMPLEX FREQ   RPGABS(47.PROGRESS..STUDENT ABSENTEEISM)

FREQUENCY / PERCENT / ROW PCT / COL PCT	IMPORTANT REASON	SOMEWHAT IMPORTA	NOT IMPORTANT	TOTAL
FREQ <= 1	29 24.17 38.16 61.70	29 24.17 38.16 78.38	18 15.00 23.68 50.00	76 63.33
FREQ >= 1	18 15.00 40.91 38.30	8 6.67 18.18 21.62	18 15.00 40.91 50.00	44 36.67
TOTAL	47 39.17	37 30.83	36 30.00	120 100.00

FREQUENCY MISSING = 1

STATISTICS FOR TABLE OF COMPLEX FREQ BY RPGABS

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.416	0.040
LIKELIHOOD RATIO CHI-SQUARE	2	6.620	0.037
MANTEL-HAENSZEL CHI-SQUARE	1	0.847	0.357
PHI		0.231	
CONTINGENCY COEFFICIENT		0.225	
CRAMER'S V		0.231	

EFFECTIVE SAMPLE SIZE = 120
FREQUENCY MISSING = 1

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 46

TABLE OF LOG USED BY XRTCNATT

LOG USED — XRTCNATT - % STUDENTS NOT ATTENTIVE AND NOT BEHAVIORAL PROBLEMS

FREQUENCY / PERCENT / ROW PCT / COL PCT	NO STUDENTS	5<= %	TOTAL
USED <= 1	6 5.36 18.75 14.29	26 23.21 81.25 37.14	32 28.57
1 < USED <= 2	20 17.86 37.04 47.62	34 30.36 62.96 48.57	54 48.21
2 < USED	16 14.29 61.54 38.10	10 8.93 38.46 14.29	26 23.21
TOTAL	42 37.50	70 62.50	112 100.00

FREQUENCY MISSING = 4

STATISTICS FOR TABLE OF LOG USED BY XRTCNATT

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	11.215	0.004
LIKELIHOOD RATIO CHI-SQUARE	2	11.470	0.003
MANTEL-HAENSZEL CHI-SQUARE	1	11.001	0.001
PHI		0.316	
CONTINGENCY COEFFICIENT		0.302	
CRAMER'S V		0.316	

EFFECTIVE SAMPLE SIZE = 112
FREQUENCY MISSING = 4

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 47

TABLE OF LOG FREQ BY RECHNG

LOG FREQ   RECHNG(66.RATING..CHANGE ACTIVITY IF NO ATTN)

FREQUENCY / PERCENT / ROW PCT / COL PCT	OF LITTLE OR NO	SOME IMPORTANCE	MAJOR IMPORTANCE	AMONG HIGHEST	TOTAL
FREQ <= 1	9	27	33	13	82
	7.89	23.68	28.95	11.40	71.93
	10.98	32.93	40.24	15.85	
	45.00	75.00	78.57	81.25	
FREQ >= 1	11	9	9	3	32
	9.65	7.89	7.89	2.63	28.07
	34.38	28.13	28.13	9.38	
	55.00	25.00	21.43	18.75	
TOTAL	20	36	42	16	114
	17.54	31.58	36.84	14.04	100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG FREQ BY RECHNG

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	3	8.958	0.030
LIKELIHOOD RATIO CHI-SQUARE	3	8.243	0.041
MANTEL-HAENSZEL CHI-SQUARE	1	6.086	0.014
PHI		0.280	
CONTINGENCY COEFFICIENT		0.270	
CRAMER'S V		0.280	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 48

TABLE OF LOG FREQ BY REFEED

LOG FREQ REFEED(74.RATING..FREQUENT INDIVIDUAL FEEDBACK)

FREQUENCY / PERCENT / ROW PCT / COL PCT	SOME IMPORTANCE	MAJOR IMPORTANCE	AMONG HIGHEST	TOTAL
FREQ <= 1	9 7.89 10.98 69.23	35 30.70 42.68 62.50	38 33.33 46.34 84.44	82 71.93
FREQ >= 1	4 3.51 12.50 30.77	21 18.42 65.63 37.50	7 6.14 21.88 15.56	32 28.07
TOTAL	13 11.40	56 49.12	45 39.47	114 100.00

FREQUENCY MISSING = 2

STATISTICS FOR TABLE OF LOG FREQ BY REFEED

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	6.004	0.050
LIKELIHOOD RATIO CHI-SQUARE	2	6.300	0.043
MANTEL-HAENSZEL CHI-SQUARE	1	3.584	0.058
PHI		0.229	
CONTINGENCY COEFFICIENT		0.224	
CRAMER'S V		0.229	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 2

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 49
TABLE OF COMPLEX FREQ BY RESAYGD

COMPLEX FREQ RESAYGD(89.RATING..SAY SOMETHING GOOD ABOUT ANS)

FREQUENCY / PERCENT / ROW PCT / COL PCT	OF LITTLE OR NO	SOME IMPORTANCE	MAJOR IMPORTANCE	AMONG HIGHEST	TOTAL
FREQ <= 1	7 5.93 9.33 87.50	26 22.03 34.67 57.78	35 29.66 46.67 72.92	7 5.93 9.33 41.18	75 63.56
FREQ >= 1	1 0.85 2.33 12.50	19 16.10 44.19 42.22	13 11.02 30.23 27.08	10 8.47 23.26 58.82	43 36.44
TOTAL	8 6.78	45 38.14	48 40.68	17 14.41	118 100.00

FREQUENCY MISSING = 3

STATISTICS FOR TABLE OF COMPLEX FREQ BY RESAYGD

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	3	8.121	0.044
LIKELIHOOD RATIO CHI-SQUARE	3	8.370	0.039
MANTEL-HAENSZEL CHI-SQUARE	1	1.398	0.237
PHI		0.262	
CONTINGENCY COEFFICIENT		0.254	
CRAMER'S V		0.262	

EFFECTIVE SAMPLE SIZE = 118
FREQUENCY MISSING = 3

CONTENT REPRESENTATION: TECHNICAL APPENDIX
TABLE 50

TABLE OF COMPLEX FREQ BY SDAYSYR

COMPLEX FREQ SDAYSYR

FREQUENCY / PERCENT / ROW PCT / COL PCT	< 180	180	180<	TOTAL
FREQ <= 1	12	42	17	71
	10.53	36.84	14.91	62.28
	16.90	59.15	23.94	
	41.38	68.85	70.83	
FREQ >= 1	17	19	7	43
	14.91	16.67	6.14	37.72
	39.53	44.19	16.28	
	58.62	31.15	29.17	
TOTAL	29	61	24	114
	25.44	53.51	21.05	100.00

FREQUENCY MISSING = 7

STATISTICS FOR TABLE OF COMPLEX FREQ BY SDAYSYR

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	2	7.262	0.026
LIKELIHOOD RATIO CHI-SQUARE	2	7.105	0.029
MANTEL-HAENSZEL CHI-SQUARE	1	7 190	0.007
PHI		0.252	
CONTINGENCY COEFFICIENT		0.245	
CRAMER'S V		0.252	

EFFECTIVE SAMPLE SIZE = 114
FREQUENCY MISSING = 7

CONTENT REPRESENTATION: TECHNICAL APPENDIX

TABLE 51

TABLE OF COMPLEX USED BY SN015 - OVERALL CURRICULUM

COMPLEX USED   SN015 - OVERALL CURRICULUM

FREQUENCY PERCENT ROW PCT COL PCT	COMPREHENSIVE GENERAL 1	CORE GEN AND SPEC COURSES 2	STREAMING BY STU INTEREST 4	TOTAL
USED <= 1	8	4	20	32
	7.14	3.57	17.86	28.57
	25.00	12.50	62.50	
	32.00	12.12	37.04	
1 < USED <= 2	4	5	15	24
	3.57	4.46	13.39	21.43
	16.67	20.83	62.50	
	16.00	15.15	27.78	
2 < USED	13	24	19	56
	11.61	21.43	16.96	50.00
	23.21	42.86	33.93	
	52.00	72.73	35.19	
TOTAL	25	33	54	112
	22.32	29.46	48.21	100.00

FREQUENCY MISSING = 9

STATISTICS FOR TABLE OF COMPLEX USED BY SN015

STATISTIC	DF	VALUE	PROB
CHI-SQUARE	4	12.349	0.015
LIKELIHOOD RATIO CHI-SQUARE	4	12.968	0.011
MANTEL-HAENSZEL CHI-SQUARE	1	4.585	0.032
PHI		0.332	
CONTINGENCY COEFFICIENT		0.315	
CRAMER'S V		0.235	

EFFECTIVE SAMPLE SIZE = 112
FREQUENCY MISSING = 9

Appendix 2

SIMS International and National Reports

NATIONAL REPORTS ON SECOND INTERNATIONAL MATHEMATICS STUDY

(NOVEMBER, 1989)

CANADA (British Columbia)

Robitaille, David F., J. Thomas O'Shea and Michael Dirks (1982) The Teaching and Learning of Mathematics in British Columbia. Victoria, BC: Ministry of Education, Learning Assessment Branch.

Robitaille, D.F. (1985) An Analysis of Selected Achievement Data from the Second International Mathematics Study. Victoria,BC: Ministry of Education, Student Assessment Branch.

CANADA (Ontario)

McLean, L., D. Raphael, M. Wahlstrom (1986) Intentions and Attainments in the Teaching and Learning of Mathematics. Reports on the Second International Mathematics Study ' Ontario. Toronto: Ontario Ministry of Education.

McLean, L., R. Wolfe, M. Wahlstrom (1987) Learning About Teaching from Comparative Studies: Ontario Mathematics in International Perspective. Toronto,Ont: Ontario Ministry of Education.

Raphael, D., M. Wahlstrom, L. McLean (1983) The Second International Study of Mathematics: An Overview of the Ontario Grade 8 Study. Toronto, Ont.: Ontario Institute for Studies in Education.

Raphael, D., M. Wahlstrom, L. McLean (1983) The Second International Study of Mathematics: An Overview of the Ontario Grade 12/13 Study. Toronto, Ont.: Ontario Institute for Studies in Education.

ENGLAND AND WALES

Cresswell, M. J.Grubb (1987) The Second International Mathematics Study in England and Wales. International Studies in Pupil Performance. Windsor, Berks: NFER-Nelson.

FINLAND

Kangasniemi, E (1988) Opetussuunnitelma ja matematijkan Koulusaavutukset. (Curriculum and student achievment in mathematics). Research Reports: Publication series A. Institute for Educational Research, University of Jyvaskyla, Finland.

FRANCE

Robin, D., E. Barrier (1985). Enquete internationale sur l'enseignement des mathematiques: Le cas francais (International Mathematics Study: The French Case)

Tome 1. INPR, Collection. National de Recherche Pedagogique. "Rapports de recherches" 8. Paris.

HONG KONG

Brimer, A., P. Griffin (1985) Mathematics Achievement in Hong Kong Secondary Schools. Center for Asian Studies, University of Hong Kong, Hong Kong.

ISRAEL

Padua, M., R. Raz, Eds. (1982) Cross-National Study in Mathematics Grades 8, 12: The Research Instruments. Israel Curriculum Center, Ministry of Education and Culture. Jerusalem. [In Hebrew.]

Lewy A. (1983) Attitudes and Attainments in Mathematics: A Technical Report of a Survey in Grades 8 and 12. Israel Curriculum Center, Ministry of Education and Culture. Jerusalem. [In Hebrew with English summary.]

Padua, H. (1983) Report on Mathematics Achievement Survey, Grades 8 and 12. Israel Curriculum Center, Ministry of Education and Culture, Jerusalem. [In Hebrew with English summary.]

Lewy, A. (1984) Mathematics achievement in Grade 12. In A. M. Mayer and P. Tamir (Eds.), Science Teaching in Israel: Origins, Development and Achievements. The Amos De-Shalit Science Teaching Center, Jerusalem. [Hebrew with English summary.]

JAPAN

Sawada, T., Ed. 1981. Mathematics Achievement of Secondary School Students. National Institute for Educational Research, Tokyo. (In Japanese)

Sawada, T., Ed. 1982. Mathematics Achievement and Associated Factors of Secondary School Students. National Institute for Educational Research, Tokyo. (In Japanese)

Sawada, T., Ed. 1983. Mathematics Achievement and Teaching in Lower Secondary School - - Grade 7. National Institute for Educational Research, Tokyo. (In Japanese)

NETHERLANDS

Kuper, J., W. J. Pelgrum, 1983. Het IEA Iweeda Wiskunda Project:Aspecten van Meetkunda (The IEA Second Mathematics Study: Aspects of Geometry). Universiteit Twente, Toegepaste Onderwijskuunde, Enschede.

Pelgrum, W.J., Th. J.H.M. Eggen, Tj Plomp, 1983. Het IEA Iweeda Wiskunda Project: Beschrijving van uitkomsten (The IEA Second Mathematics Study: Description of Results). Universiteit Twente, Toegepaste Onderwijskuunde, Enschede.

Pelgrum, W.J., Th. J.H.M. Eggen, Tj Plomp, 1983. Het IEA Iweeda Wiskunda Project: Opzet en uitvoering (The IEA Second Mathematics Study: Design and Execution). Universiteit Twente, Toegepaste Onderwijskuunde, Enschede.

Pelgrum, W.J., Th. J.H.M. Eggen, Tj Plomp, 1983. Het IEA Iweeda Wiskunda Project: Analyse van uitkomsten (The IEA Second Mathematics Study: Analysis of Results). Universiteit Twente, Toegepaste Onderwijskuunde, Enschede.

NEW ZEALAND

Mathematics Achievement in New Zealand Secondary Schools 1987. Department of Education, Wellington, New Zealand.

SWEDEN

Ministry of Education 1986 Matematik I Skolan: oversyn av undervisningen i matematik inom skclvasendet (Mathematics in the School. Examining the Teaching of Mathematics in the the School System). Ministry of Education and Liber AllmannaForlaget, Stockholm, Sweden.

Ministry of Education 1986 Matematik for Alla: ett diskussionsmateiral [Mathematics for All: Material for Discussion]. Ministry of Education and Liber Utbilningsforlget, Stockholm, Sweden.

Ministry of Education 1987 Mamnaren. 86/87 (2-3): 1-12B.

Murray, R. R. Liljefors 1983 Matematik i Svensk Skola. Skoloverstyrelsen.

THAILAND

Thai National Committee for the Second International Mathematics Study 1984 The Analysis of the Mathematics Curriculum. The Institute for the Promotion of Teaching Science and Technology, Bangkok. [In Thai]

Thai National Committee for the Second International Mathematics Study 1985. The Evaluation of Mathematics Teaching and Learning in Mattayom Suksa 5 Classes. The Institute for the Promotion of Teaching Science and Technology, Bangkok.[In Thai]

Thai National Committee for the Second International Mathematics Study. 1986. The Evaluation of Mathematics Teaching and Learning in Mattayom Suksa 2 Classes. The Institute for the Promotion of Teaching Science and Technology, Bangkok. [In Thai]

UNITED STATES

Chang, A. et. al., 1985. Technical Report I: Item Level Achievement and OTL Data. Stipes Publishing Company, Champaign, IL.

Crosswhite, J. et.al., 1985. United States Summary Report: Second International Mathematics Study. Stipes Publishing Company, Champaign, IL.

Crosswhite, J. et. al., 1986. Detailed National Report. Stipes Publishing Company, Champaign, IL.

McKnight, C. et al. 1987. The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective. Stipes Publishing Company, Champaign, IL.

INTERNATIONAL REPORTS ON THE SECOND INTERNATIONAL MATHEMATICS STUDY

Garden, R.A., 1987. Second IEA Mathematics Study Sampling Report. Center for Education Statistics, U.S. Department of Education,Washington, D.C., USA.

Technical Reports

Dossey, J.A. et al. 1987. Mathematics teachers in the United States and other countries participating in the Second International Mathematics Study. University of Illinois at Urbana-Champaign, Urbana, IL.

Jajji, G. 1986. The used of calculators and computers in mathematics classes in twenty countries: Summary Report. University of Illinois at Urbana-Champaign, Urbana, IL.

Kifer, E., R.R. Wolfe, W. H. Schmidt, 1985. Cognitive Growth in Eight Countries: Lower Secondary School. University of Illinois at Urbana-Champaign, Urbana, IL.

Livingstone, I.D., 1986 Perceptions of the Intended and Implemented Mathematics Curriculum. Center for Education Satistics. U.S. Department of Education, Washington, D.C., USA.

Miller, D.M., R.L. Linn, 1986. Cross-national achievement with differential retention rates. University of Illinois at Urbana-Champaign, Urbana, IL.

Oldham, E. E., 1986. Qualifications of mathematics teachers. Trinity College, Dublin, Ireland.

Pelgrum, W.J., Th.J.H.M. Eggan, T.J. Plomp, 1985. The implemented and attained mathematics curriculum: A comparison of twenty countries. Twente University of Technology, Enschede, The Netherlands.

Robitaille, D.F., A.R. Taylor, 1985. A comparative review of students' achievement in the First and Second IEA Mathematics Studies. University of British Columbia, Vancouver, Canada.

Wolfe, R.G. 1987. A comparative analysis across eight countries of growth in mathematics achievement: The effects of one year of instruction. Ontario Institute for Studies in Education, Toronto, Canada.

Wolfe, R.G. 1986. Training Manual for Use of the Data Bank of the Longitudinal, Classroom Process Surveys for Population A in the IEA Second International Mathematics Study. Ontario Institute for Studies in Education, Toronto, Canada.

Wolfe, R.G., D. Raphael, 1985. Reports of Homework Assignment Around the World: Results from the Second International Mathematics Study. Ontario Institute for Studies in Education, Toronto, Canada.

Appendix 16

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991